

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: May 13, 2004, 08:21:22 ; Search time 67.5 Seconds
(without alignments)
10919.267 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 2048
Sequence: 1 ataaatcacgtcattgacatt.....aagccgttaaaagtgccgac 1168

Scoring table:
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO spoel_p/US10030529/runat 13052004 081452 2610/app query.fasta 1.1351
-DB=SPTRMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40 cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10030529 @CGN 1.1 86 @runat 13052004 081452 2610 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL 25:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1342	65.5	257	16	Q9K2H6 haemophilus

C	2	1299.5	63.5	264	2	Q9K3B8	
	3	1289	62.9	271	2	Q9L916	
	4	1283	62.6	273	2	Q9L917	
	5	1262	61.6	248	2	Q9L919	
	6	1205	58.8	263	2	Q9L918	
	7	235	11.8	440	16	Q7VN22	
	8	225	11.0	295	2	Q8RR69	
	9	225	11.0	295	2	Q8RO63	
	10	222	10.8	504	2	Q9LA56	
	11	222	10.8	511	9	Q9MC18	
	12	220	10.7	392	2	Q9LA60	
	13	217	10.6	355	2	Q8KUJ1	
	14	217	10.6	362	2	Q8KH7	
	15	217	10.6	364	16	Q9JXK7	
	16	217	10.6	459	2	Q8VW24	
	17	209	10.2	391	2	Q8KUJ6	
	18	209	10.2	398	2	Q8KI42	
	19	208.5	10.2	487	2	Q9LA53	
	20	204	10.0	405	2	Q8KHP5	
	21	204	10.0	405	2	Q8KH85	
	22	198	9.7	668	2	Q8GH86	
	23	196	9.6	630	2	Q9XD55	
	24	193	9.4	613	2	Q9XD53	
	25	193	9.4	616	2	Q848S2	
	26	193	9.4	877	2	Q848S1	
27	193	9.4	889	2	Q9L961		
28	193	9.4	894	2	Q9L962		
29	190.5	9.3	576	2	O54407		
30	190.5	9.3	674	2	Q9XD51		
31	190.5	9.3	684	2	Q9L963		
32	190.5	9.3	686	2	Q8RTB2		
33	186	9.1	422	2	Q93KR4		
34	186	9.1	422	2	Q56930		
35	180.5	8.8	422	2	Q84GR6		
36	161	7.9	454	2	O85267		
37	159.5	7.8	2712	16	Q9F3X5		
C	38	151	7.6	419	16	Q9CLN4	
	39	132	6.4	155	16	Q8YJ81	
	40	130	6.3	278	16	Q8FYM1	
	41	130	6.5	420	16	Q8FEZ1	
	42	129	6.5	227	16	Q7UBW0	
	43	129	6.5	399	16	Q83K08	
	44	129	6.5	413	16	Q8XFY3	
	45	129	6.5	428	16	Q8X9C3	
		Q9K3b8	haemophilus				
		Q9L916	haemophilus				
		Q9L917	haemophilus				
		Q9L919	haemophilus				
		Q9L918	haemophilus				
		Q7vn22	haemophilus				
		Q8rr69	actinobacil				
		Q8rq63	actinobacil				
		Q9la56	escherichia				
		Q9mc18	bacteriopha				
		Q9la60	escherichia				
		Q8kuj1	neisseria m				
		Q8khf7	neisseria m				
		Q9jxk7	neisseria m				
		Q8vw24	escherichia				
		Q8kuj6	neisseria m				
		Q8ki42	neisseria m				
Q9la53		escherichia					
Q8knp5		neisseria m					
Q8kh85		neisseria m					
Q8gh86		moraxella c					
Q9xd55		moraxella c					
Q9xd53		moraxella c					
Q848s2		moraxella c					
Q848s1		moraxella c					
Q9l961	moraxella c						
Q9l962	moraxella c						
O54407	moraxella c						
Q9xd51	moraxella c						
Q9l963	moraxella c						
Q8rtb2	moraxella c						
Q93kr4	yersinia en						
Q56930	yersinia en						
Q84gr6	yersinia en						
O85267	yersinia en						
Q9f3x5	pasteurella						
Q9cln4	pasteurella						
Q8yj81	brucella me						
Q8fym1	brucella su						
Q8fez1	escherichia						
Q7ubw0	shigella fl						
Q83k08	shigella fl						
Q8xfy3	salmonella						
Q8x9c3	escherichia						

ALIGNMENTS

RESULT 1	
Q9K2H6	PRELIMINARY; PRT; 257 AA.
ID	Q9K2H6
AC	Q9K2H6; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Serum resistance protein DsrA (Serum resistance protein DsrA).
GN	DSRA OR HD0769.
OS	Haemophilus ducreyi.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC	Pasteurellaceae; Haemophilus.
OX	NCBI_TaxID=730;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CIP542(CDC), 35000, and 406;
RX	MEDLINE=20143779; PubMed=10678980;
RA	Elkins C., Morrow K.J. Jr., Olsen B.;
RT	"Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
RL	Protein DsrA.";
RL	Infect. Immun. 68:1608-1619(2000).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=35000HP / ATCC 700724;
RA	Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,

RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RL "The complete genome sequence of Haemophilus ducreyi."
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF197007; AAF37813.1; -;
DR EMBL; AF197007; AAF37807.1; -;
DR EMBL; AF187005; AAF37811.1; -;
DR EMBL; AE017152; AAP95674.1; -;
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28524 MW; F497BF2CD5666938 CRC64;

Alignment Scores:
Pred. No.: 2.93e-104 Length: 257
Score: 1342.00 Matches: 257
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.53% Indels: 0
DB: 16 Gaps: 0

US-10-030-529A-1 (1-1168) x Q9K2H6 (1-257)

QY 101 ATGAAATTAATGTTAGTTCGCCGTAGTGGATTAGCTTGTCTACTATTACAAATG 160
DB 1 MetLysileLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTCTGGAGTATCTTTGTATAGCTATGACTATGACTAT 220
DB 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACATGCTCTAATCAAGCGGTTTCGATATTAAAGTCCAGGATT 280
DB 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGATGATTTCTAAACAGGCTACTTATCTTGAATTACAGATTAT 340
DB 61 LysMetLysProLysGluTyrIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCTTATATCTCTGTCGATACATGCTCTCGGTTTCCTAGCCCTATAGT 400
DB 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
QY 401 TTATATCCGATGTCCTGATCAACTTGAATTAATCGCAGCAGCTGAAATTGAAT 460
DB 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTATTTTAAACGATTAAGACACGATTTTAAATTAAGTTCTTGATGCAGT 520
DB 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
QY 521 ATTTCCAAATTAACAAATATTGATATAGTAAATTTTACTAGACTCGGTACT 580
DB 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTAGATGATTTCTTATCGTATGATGAACAAATACACATATCAATTAAGTTGTCT 640
DB 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
QY 641 AAAGAATTGCAAACTGGTTTACCCAAACCAATCAGCATTTGTCTATGTTAGTCAACCAAT 700
DB 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
QY 701 GGTGTAGCAAAACGAGGTTTCTGCTGGGTAGGAGGTTATAGATAAACTGCATTA 760
DB 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
QY 761 GCATTTGGTGTCCGCTCAGCATTTACTGATCCGTTTACCGCTTAAAGCGGTGTAGCGTTC 820
DB 221 AlalleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
QY 821 AATACCTACAATGGCGGATGCTTATGCTGCTTCTGTTGGTATGAATTC 871
DB 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257

RESULT 2
Q9K3B8 PRELIMINARY; PRT; 264 AA.
ID Q9K3B8
AC Q9K3B8 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein Dera.
GN DSEA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIPA77, and CIPA75;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
Protein Dera.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187003; AAF37809.1; -;
DR EMBL; AF187002; AAF37808.1; -;
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 264 AA; 29444 MW; 11AF6C124D7BD126 CRC64;

Alignment Scores:
Pred. No.: 1.09e-100 Length: 264
Score: 1299.50 Matches: 253
Percent Similarity: 95.83% Conservative: 0
Best Local Similarity: 95.83% Mismatches: 4
Query Match: 63.45% Indels: 7
DB: 2 Gaps: 1

US-10-030-529A-1 (1-1168) x Q9K3B8 (1-264)

QY 101 ATGAAATTAATGTTAGTTCGCCGTAGTGGATTAGCTTGTCTACTATTACAAATG 160
DB 1 MetLysileLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTCTGGAGTATCTTTGTATAGCTATGACTATGACTAT 220
DB 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACATGCTCTAATCAAGCGGTTTCGATATTAAAGTCCAGGATT 280
DB 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGATGATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTA 340
DB 61 LysMetLysProLysGluTyrIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCTTATATCTCTGTCGATACATATGCTCTCGGCTTCTCTAGCCCTATAGT 400
DB 81 MetProTyrThrProValLeuValThrTyrAlaHisAspValProProSerSerIleLeu 100
QY 401 TTATATCCGATGTCCTGATCAACTTGAATTAATCGCAGCAGCTGAAATTGAAT 460
DB 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTATTTTAAACGATTTAAGCACCATTTTAAATTAAGTTCTTGATGCAGT 520
DB 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
QY 521 ATTTCCAAATTAACAAATATTGATACTATAAGTAAATTTTACTAGAACTGGGTACT 580
DB 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTAGATGATTTCTTATCGTATGATGAACAA-----NATACA 619
DB 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180

```

Qy 620 CATAATCAATAAGTTCTCTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCATG 679
Db 181 HisanilleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 200
Qy 680 TCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTTCTGCTGCGGTAGGAGGT 739
Db 201 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaValGlyGly 220
Qy 740 TATAGATATAAAGTGCATTAGCCATTGGTGTGGCTCGCTCACCATTGATCGCTTTACC 799
Db 221 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 240
Qy 800 GCTAAAGCGGTAGCGTTCAATACCTACCAATGCGCGCATGCTTATGGTGTCTTGTT 859
Db 241 AlalysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 260
Qy 860 GGTTATGAATTC 871
Db 261 GlyTyrGluPhe 264

RESULT 3
Q9L916 PRELIMINARY; PRT; 271 AA.
AC Q9L916;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187009; AAF37815.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 271 AA; 30091 MW; 5A9DC55C6DA6EF1C CRC64;

Alignment Scores:
Pred. No.: 8.28e-100 Length: 271
Score: 1289.00 Matches: 252
Percent Similarity: 92.99% Conservative: 0
Best Local Similarity: 92.99% Mismatches: 5
Query Match: 62.94% Indels: 14
DB: 2 Gaps: 1

US-10-030-529a-1 (1-1168) x Q9L916 (1-271)
Qy 101 ATGAAAATTAAATGTTAGTTCGCGTAGTGGATTAGCTTCTTACTATTACAAACATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Qy 161 GCTCAGACGCGCAAGATTGCTGGAGTATCTCTTTGTATAGCTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Qy 221 GGTAAAGGTAATGACTTGTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGCGATT 280
Db 41 GlyLysGlyLysIleThrTrpSerAsnGlnGlyGlyPheAspIleLysValProGlyIle 60
Qy 281 AAAATGAAGCAAAAGAAATGATTCTTAAACAGCGCTACTTCTTGAATACAGCAATTAT 340
Db 61 LysMetLysProlLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80

```

```

Qy 341 ATGCTTTATACTCTGTTCTCGTGACATATGCTCTCGCGGTTTCTCTAGCCCTATACGT 400
Db 81 MetProTyrThrProValLeuValThrSerAlaProAspValProProSerSerIleLeu 100
Qy 401 TTATATCCGATGCTGATCTCTGATCAACTTGGATAAATCGGCAGCAGCTGAAATTGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
Qy 461 TTGTATAGTATTATTAACGATTTAAGACACAGATTTAAATTAAGTTCTTGATGCAGGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
Qy 521 ATTTCCAAAAATAACAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
Qy 581 TATTAGATGATCTTATCGTATGATGAAACA----- 613
Db 161 TyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180
Qy 614 -----AATACACATAATATCAATAAGTTGTCTAAAGATTGCAAACTGGT 658
Db 181 HisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGlnThrGly 200
Qy 659 TTAGCCAAACCAATCAGCATTTCTGTTAGTGTCAACCAATGTTGTAGGCAAAACGAGC 718
Db 201 LeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLysThrSer 220
Qy 719 GTTTCTGCTCGGTAGGAGTTATAGAGATAAAACTGCATTAGCCATTGGTGTGCGGTCA 778
Db 221 ValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeuAlaIleGlyValGlySer 240
Qy 779 CGCATTTACTGCTTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTCAATGCGCGC 838
Db 241 ArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGly 260
Qy 839 ATGCTTTATGCTCTCTGTTGGTTATGAATTC 871
Db 261 MetSerTyrGlyAlaSerValGlyTyrGluPhe 271

RESULT 4
Q9L917 PRELIMINARY; PRT; 273 AA.
AC Q9L917;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187008; AAF37814.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 273 AA; 30281 MW; 8C620D6B5544E3EE CRC64;

Alignment Scores:
Pred. No.: 2.64e-99 Length: 273
Score: 1283.00 Matches: 253
Percent Similarity: 92.67% Conservative: 0
Best Local Similarity: 92.67% Mismatches: 4
Query Match: 62.65% Indels: 16
DB: 2 Gaps: 2

```

```
US-10-030-529A-1 (1-1168) x Q9L919 (1-273)
QY 101 ATGAAATTAATGTTAGTTCGGGTAGTGGATTCCTTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTGCTGGAGTATCTTTGTATAGTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACTTGTCTTAATCAAGCGGTTTCGATATTAAGTCCAGGGATT 280
Db 41 GlyLysGlyLysTyrThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGAATGGATTCTTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCCCTTATCTCTGTCGATATGCTCTCGGCTTCTCCTAGC-----CCT 394
Db 81 MetProTyrThrProValLeuValThrSerAlaProAspValSerProSerSerIleSer 100
QY 395 ATACTGTTATATCCGATCTCTGATCAACTTGAATAAATCGGCAGCAGCTGAAA 454
Db 101 IleLeuLeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLys 120
QY 455 TTGAATTTGTATAGTATTTTAAACGATTTAGACAGATTTTAAATTAAGTTCTTGAT 514
Db 121 LeuAsnLeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAsp 140
QY 515 GCACGTTATCCAAAATAACAAATATGATAGTATGATAGTAAATATTTACTAGAACTG 574
Db 141 AlaArgIleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeu 160
QY 575 GGTACTTATTTAGATTTCTTATCGTATGATGGAACAA----- 613
Db 161 GlyThrTyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLys 180
QY 614 -----AATACATAATATCAATAGTTGTCTTAAGAATTGCAA 652
Db 181 AsnThrHisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGln 200
QY 653 ACTGGTTTAGCCAAACATCAGCATTTCTATGTTAGTGAACCAAAATGGTGTAGCAAA 712
Db 201 ThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLys 220
QY 713 ACGAGCGTTTCTGCTGCGGTAGAGGTTATAGAGATAAACTGCAATTAGCCATTGGTCTC 772
Db 221 ThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeuAlaIleGlyVal 240
QY 773 GCCTCAGCATTTACTGATCGCTTTACCGTAAAGCGGTGTAGCGTTCAATACCTACAAT 832
Db 241 GlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsn 260
QY 833 GSCGCGCTCTTATGGTCTTCTGTTGTTATGAATTC 871
Db 261 GlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 273

RESULT 5
Q9L919
ID Q9L919 PRELIMINARY; PRT; 248 AA.
AC Q9L919,
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN (1)
```

RP SEQUENCE FROM N.A.
RC STRAIN=CIP542(Can);
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
RL Protein DsrA.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187004; AAF37810.1; -.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
SQ SEQUENCE 248 AA; 27815 MW; 245C809F24E6A815 CRC64;

Alignment Scores:
Pred. No.: 1.52e-97 Length: 248
Score: 1262.00 Matches: 242
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 61.62% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x Q9L919 (1-248)

QY 101 ATGAAATTAATGTTAGTTCGGGTAGTGGATTCCTTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTGCTGGAGTATCTTTGTATAGTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACTTGTCTTAATCAAGCGGTTTCGATATTAAGTCCAGGGATT 280
Db 41 GlyLysGlyLysTyrThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGAATGGATTCTTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCCCTTATCTCTGTCGATATGCTCTCGGCTTCTCCTAGCCCTACTATG 400
Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
QY 401 TTATATCCGATGTCCTGATCCCTGATCAACTTGAATAAATCGGCAGCAGCTGAAATTTGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTATTTTAAAGATTTAAGACACAGATTTTAAATTAAGTTCTTGTATCCAGCT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
QY 521 ATTTCCAAAATAACAAAATATTGATATAAGTAAATATTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTTAGATGATTTCTTATCGTATGATGGAACAAAATACACATAATATCAATAGTTGTCT 640
Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
QY 641 AAAGAATTGCAACTGGTTTAGCCAAACCAATCAGCATTTGTCTATGTTAGTGAACCAAT 700
Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
QY 701 GGTGTAGCAAAACGAGCGTTTCTGCTGGGTAGGAGGTATAGAGATAAATCTCATTA 760
Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
QY 761 GCCATTGTCGCGCTCACGCATTACTGATCGCTTTACCGCTTAAAGCGGTGTAGCGTTTC 820
Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
QY 821 AATACCTTAC 829
Db 241 AsnThrPhe 243


```

RESULT 6
Q9L918      PRELIMINARY;          PRT; 263 AA.
ID  Q9L918;
AC  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE  Serum resistance protein Dera.
GN  DSPA.
OS  Haemophilus ducreyi.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC  Pasteurellaceae; Haemophilus.
OX  NCBI_TaxID=730;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CHIA;
RX  MEDLINE=20143779; PubMed=10678980;
RA  Elkins C., Morrow K.J. Jr., Olsen B.;
RT  "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
RT  Protein Dera.";
RL  Infect. Immun. 68:1608-1619 (2000).
DR  EMBL; AF187006; AAF37812.1; -.
DR  InterPro; IPR005594; Yada.
DR  Pfam; PF03895; Yada; 1.
SQ  SEQUENCE 263 AA; 29170 MW; 858B91BF765BE8F1 CRC64;

Alignment Scores:
Pred. NO.:      9.32e-93      Length:      263
Score:          1205.00      Matches:    235
Percent Similarity: 92.05%    Conservative: 8
Best Local Similarity: 89.02% Mismatches: 13
Query Match:     58.84%      Indels:     8
DB:              2          Gaps:       2

US-10-030-529A-1 (1-1168) x Q9L918 (1-263)
QY  101 ATGAATAATTAATGTTAGTTCGCGTAGTGGATTAGCTGCTTCTACTATTACAAATG 160
Db  |||||
QY  1 MetLysIleuYsCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Db  |||||
QY  161 GCTCAGCAGCGCCCAAGATTTGCTGGAGTACTCTTTGTATAGCTATGAGTATGACTAT 220
Db  |||||
QY  21 AlaGlnGlnProProlYsPheAlaGlyValSerSerLeuAspSerTyrrGlutyrAspTyrr 40
Db  |||||
QY  221 GGTAAAGGTAATAGGACTTGCTTAATGAAGCGGTTTCGATATATAAGTCCAGGAT 280
Db  |||||
QY  41 GlyLysGlyLysTrpThrTrpSerGluLysAspGlyPheAspIleLysAlaProGlyIle 60
Db  |||||
QY  281 AAAATGAGCCCAAGATGATTTCTAAACAGGCTACTTATCTTGAATTACAGCATAT 340
Db  |||||
QY  61 LysMetLysProlYsLysTrpIleSerArgGlnAlaThrTyrrLeuGlyLeuGlnHisTyrr 80
Db  |||||
QY  341 ATGCTTATATCTCTGTTCTCGTGACATATCTCTCGCGGTTCTCTAGCCCTATATCTG 400
Db  |||||
QY  81 MetProlyrrThrProValLeuValThrTyrrAla---SerAlaGluProAsnThrValLeu 99
Db  |||||
QY  401 TTATATCCGATGCTCGATCACTTGGATAAATCGGCAGCAGCTGAAATTTGAAT 460
Db  |||||
QY  100 LeuTyrrProMetProAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 119
Db  |||||
QY  461 TTGTATAGTTATTTAAGCATTTAAGACACGATTTTAAATTAAGTTCTTGTGTCAGCT 520
Db  |||||
QY  120 LeuTyrrSerTyrrPheAsnAspLeuArgHisGlyPheLysLeuAsnValLeuAspAlaArg 139
Db  |||||
QY  521 ATTTCCCAATAACAAATATTGATCTACTAATAATATTACTAGACTGGTACT 580
Db  |||||
QY  140 IleserGlnAsnLysGlnAsnIleAspThrIleSerGluTyrrLeuLeuLysLeuGlyThr 159
Db  |||||
QY  581 TATTATAGATGTTCTTATCGTATGATGAACAA-----AATACA 619
Db  |||||
QY  160 TyrrLeuAspSerSerTyrrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 179
Db  |||||

RESULT 7
Q7VN22      PRELIMINARY;          PRT; 440 AA.
ID  Q7VN22;
AC  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hypothetical protein.
GN  HD0770.
OS  Haemophilus ducreyi.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC  Pasteurellaceae; Haemophilus.
OX  NCBI_TaxID=730;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=35000HP / ATCC 700724;
RA  Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA  Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT  "The complete genome sequence of Haemophilus ducreyi.";
RL  Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE017152; AAP95675.1; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 440 AA; 49472 MW; AF37BC89E59ABABA CRC64;

Alignment Scores:
Pred. NO.:      2.67e-11      Length:      440
Score:          235.00      Matches:    48
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     11.78%      Indels:     0
DB:              16          Gaps:       0

US-10-030-529A-1 (1-1168) x Q7VN22 (1-440)
QY  1168 GTCCGACCTTTACGCTTAATTTAGACATTTAGAAAATTCGATAGAGGTACC 1109
Db  |||||
QY  393 ValArgThrPheAsnGlyLeuIleLeuGluHisLeuGluLysIleProIleGluGlyThr 412
Db  |||||
QY  1108 CAATTTAGGTAAACCACTTAATATTACCGTTTTAGAGTGGCGGATAATATGTTGAAA 1049
Db  |||||
QY  413 GlnPheArgLeuAsnGlnLeuAsnIleThrValLeuGluValAlaAspAsnMetValLys 432
Db  |||||
QY  1048 AAGTGAAGTAGAATGATGAA 1025
Db  |||||
QY  433 LysValLysValGluLeuIleGlu 440
Db  |||||

RESULT 8
Q8RR69      PRELIMINARY;          PRT; 295 AA.
ID  Q8RR69;
AC  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE Outer membrane protein 100.
GN OMP100.
OS Actinobacillus actinomycetemcomitans (Haemophilus
actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RA Komatsu H., Asakawa R., Kawai T., Ochiai K., Fujiwara T.,
RA Taubman M.A., Kurihara H., Sugai M.;
RT "Identification of six major outer membrane proteins from
RT Actinobacillus actinomycetemcomitans.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064943; BAB96905.1; -
DR InterPro; IPR008378; Ubiq_surface.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
DR PRINTS; PR01804; UBIQUITOUSP.
SQ SEQUENCE 295 AA; 32062 MW; C8AB93C40D7E1C69 CRC64;

Alignment Scores:
Pred. No.: 1.79e-10 Length: 295
Score: 225.00 Matches: 55
Percent Similarity: 51.45% Conservative: 34
Best Local Similarity: 31.7% Mismatches: 54
Query Match: 10.9% Indels: 30
DB: 2 Gaps: 2

US-10-030-529A-1 (1-1168) x Q8R69 (1-295)

QY 431 GGAATAATCGCAGCAGCTGAATTTGTATAGTATTTTAAAGATTAAAGACAC 490
DB 127 GlyValAsnArgValLeuGlnAsnValAspValArgSerThrGluAsnAlaAlaArgSer 146
QY 491 GATTTTAAATTAAGATCTTGTGACGATTTCCAAAAATAACAAAAATTTGAT--- 547
DB 147 -----ArgAlaAsnGluGlnLysIleAlaGluAsnLysLysAlaIleGluAsn 162
QY 547 ----- 547
DB 163 LysAlaAspLysAlaAspValGluLysAsnArgAlaAspIleAlaAlaAsnSerArgAla 182
QY 548 -----ACTATAGTAATAATTTTACTAGACCTGGGTACTTATTAGATGAT 592
DB 183 IleAlaThrPheArgSerSerGlnAsnIleAlaLeuThrThrLysValAspArg 202
QY 593 TCTTATCGTATGGAACAAATACACATAATATCAATTAAGTTCTCTAAAGAAATTCGAA 652
DB 203 AsnThrAlaArgIleAspArgLeuAspSerArgValAsnGluLeuAspLysGluValLys 222
QY 653 ACTGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAA 712
DB 223 AsnGlyLeuAlaSerGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlySer 242
QY 713 ACGACGTTTCGCTGCGTAGGAGTTATAGAGATAAACTGCAATAGCCATTGCTGTC 772
DB 243 LeuAsnLeuSerAlaAlaValGlyGlyTyrLysSerLysThrAlaLeuAlaValGlySer 262
QY 773 GGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAAAT 832
DB 263 GlyTyrArgPheAsnGlnAsnValAlaAlaLysAlaGlyValAlaValSerThrAsnGly 282
QY 833 GCGCGCATGCTTATGGTCTCTGTTGGTTATGAATTC 871
DB 283 GlySerAlaThrTyrAsnValGlyLeuAsnPheGluTrp 295

RESULT 9

Q8RQ63 PRELIMINARY; PRT; 295 AA.

AC Q8RQ63;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Putative adhesin/invasin.
GN AA32-1-2.
OS Actinobacillus actinomycetemcomitans (Haemophilus
actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RA Lepine G., Li L., Ellen R.P.;
RT "Cloning and characterization of three invasive genes of
RT Actinobacillus actinomycetemcomitans.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316502; AAL91673.1; -
DR InterPro; IPR008378; Ubiq_surface.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
DR PRINTS; PR01804; UBIQUITOUSP.
SQ SEQUENCE 295 AA; 32034 MW; C9B75DD9DD7CCDBA CRC64;

Alignment Scores:
Pred. No.: 1.79e-10 Length: 295
Score: 225.00 Matches: 55
Percent Similarity: 51.45% Conservative: 34
Best Local Similarity: 31.7% Mismatches: 54
Query Match: 10.9% Indels: 30
DB: 2 Gaps: 2

US-10-030-529A-1 (1-1168) x Q8RQ63 (1-295)

QY 431 GGAATAATCGCAGCAGCTGAATTTGTATAGTATTTTAAAGATTAAAGACAC 490
DB 127 GlyValAsnArgValLeuGlnAsnValAspValArgSerThrGluAsnAlaAlaArgSer 146
QY 491 GATTTTAAATTAAGATCTTGTGACGATTTCCAAAAATAACAAAAATTTGAT--- 547
DB 147 -----ArgAlaAsnGluGlnLysIleAlaGluAsnLysLysAlaIleGluAsn 162
QY 547 ----- 547
DB 163 LysAlaAspLysAlaAspValGluLysAsnArgAlaAspIleAlaAlaAsnSerArgAla 182
QY 548 -----ACTATAGTAATAATTTTACTAGACCTGGGTACTTATTAGATGAT 592
DB 183 IleAlaThrPheArgSerSerGlnAsnIleAlaLeuThrThrLysValAspArg 202
QY 593 TCTTATCGTATGGAACAAATACACATAATATCAATTAAGTTCTCTAAAGAAATTCGAA 652
DB 203 AsnThrAlaArgIleAspArgLeuAspSerArgValAsnGluLeuAspLysGluValLys 222
QY 653 ACTGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAA 712
DB 223 AsnGlyLeuAlaSerGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlySer 242
QY 713 ACGACGTTTCGCTGCGTAGGAGTTATAGAGATAAACTGCAATAGCCATTGCTGTC 772
DB 243 LeuAsnLeuSerAlaAlaValGlyGlyTyrLysSerLysThrAlaLeuAlaValGlySer 262
QY 773 GGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAAAT 832
DB 263 GlyTyrArgPheAsnGlnAsnValAlaAlaLysAlaGlyValAlaValSerThrAsnGly 282
QY 833 GCGCGCATGCTTATGGTCTCTGTTGGTTATGAATTC 871
DB 283 GlySerAlaThrTyrAsnValGlyLeuAsnPheGluTrp 295

RESULT 10

Q9LA56

ID Q9LA56 PRELIMINARY; PRT; 504 AA.

AC Q9LA56;

DT 01-OCT-2000 (Tremblrel. 15, Created)


```
Query Match: 10.60% Indels: 24
DB: 2 Gaps: 3

US-10-030-529A-1 (1-1168) x Q8KH7 (1-362)

QY 479 GATTAAAGACACGATTTTAAATTAAGTCTTGATGCACGTATTTCCAAAATAAACA 538
   |||:||||| ||| |||:|||||
Db 254 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 271
QY 539 AATATTGATCTAATAAGTAATATTTACTAGAACTGGGTACTTTATTTAGATGATCTTAT 598
   |||:||||| ||| |||:|||||
Db 272 ArgIleAspSerLeuAspLys----- 278
QY 599 CGTATGATGGACAAATACACATAATCAATAAGTTGTCTAAAGAAATTCCAAACTGGT 658
   |||:||||| ||| |||:|||||
Db 279 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 290
QY 659 TTAGCCCAACCAATCAGCATTTCTATGTTAGTCAACCAATGGTGTAGGCAAAACGAGC 718
   |||:||||| ||| |||:|||||
Db 291 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 310
QY 719 GTTCTGCTCGGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
   |||:||||| ||| |||:|||||
Db 311 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 330
QY 779 CGCATTACTGCTTACCGTAAAGCGGTCTAGCGTTCAATACCTCAATGCGC--- 835
   |||:||||| ||| |||:|||||
Db 331 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 350
QY 836 GGCATGTCTTATGGTGTCTGTTGGTTATGAATTC 871
Db 351 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 362

RESULT 15
Q9JXX7 PRELIMINARY; : PRT; 364 AA.
AC Q9JXX7;
DT 01-OCT-2000 (TrEMBLrel. 15; Created)
DT 01-OCT-2000 (TrEMBLrel. 15; Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24; Last annotation update)
DE Adhesin/invasin, putative.
GN NMB1994.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RA MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002548; AAF42321.1; -.
DR PIR; A81019; A81019.
DR TIGR; NMB1994; -.
DR InterPro; IPR005594; Yada.;
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 364 AA; 37972 MW; B9FF591FD25F853F CRC64;
```

Alignment Scores:
Pred. No.: 8.53e-10 Length: 364
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19

```
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 16 Gaps: 3

US-10-030-529A-1 (1-1168) x Q9JXX7 (1-364)

QY 479 GATTAAAGACACGATTTTAAATTAAGTCTTGATGCACGTATTTCCAAAATAAACA 538
   |||:||||| ||| |||:|||||
Db 256 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 273
QY 539 AATATTGATCTAATAAGTAATATTTACTAGAACTGGGTACTTTATTTAGATGATCTTAT 598
   |||:||||| ||| |||:|||||
Db 274 ArgIleAspSerLeuAspLys----- 280
QY 599 CGTATGATGGACAAATACACATAATCAATAAGTTGTCTAAAGAAATTCCAAACTGGT 658
   |||:||||| ||| |||:|||||
Db 281 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 292
QY 659 TTAGCCCAACCAATCAGCATTTCTATGTTAGTCAACCAATGGTGTAGGCAAAACGAGC 718
   |||:||||| ||| |||:|||||
Db 293 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 312
QY 719 GTTCTGCTCGGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
   |||:||||| ||| |||:|||||
Db 313 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 332
QY 779 CGCATTACTGCTTACCGTAAAGCGGTGTAGCGTTCAATACCTCAATGCGC--- 835
   |||:||||| ||| |||:|||||
Db 333 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 352
QY 836 GGCATGTCTTATGGTGTCTGTTGGTTATGAATTC 871
Db 353 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 364
```

Search completed: May 13, 2004, 08:28:31
Job time : 73.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May-13, 2004, 08:14:11 ; Search time 60 Seconds
(without alignments)
1210.245 Million cell updates/sec

Title: US-10-030-529A-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTTTM.....VAFNTYNGMGSMYGASGVGYEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqp_29Jan04:*
2: Geneseqp_1990s:*
3: Geneseqp_1990s:*
4: Geneseqp_2000s:*
5: Geneseqp_2000s:*
6: Geneseqp_2003as:*
7: Geneseqp_2003bs:*
8: Geneseqp_2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1342	100.0	257	4	AAB31713 Amino aci
2	1342	100.0	257	4	AAB31705 Amino aci
3	1342	100.0	257	4	AAB31709 Amino aci
4	1299.5	96.8	264	4	AAB31706 Amino aci
5	1299.5	96.8	264	4	AAB31707 Amino aci
6	1289	96.1	271	4	AAB31711 Amino aci
7	1283	95.6	273	4	AAB31712 Amino aci
8	1259	93.8	242	4	AAB31708 Amino aci
9	1205	89.8	263	4	AAB31710 Amino aci
10	217	16.2	355	6	ABU07925 Neisseria
11	217	16.2	357	6	ABU07926 Neisseria
12	217	16.2	362	6	ABU07915 Neisseria
13	217	16.2	364	3	AAY75736 Neisseria
14	217	16.2	364	4	AAU27562 Neisseria
15	217	16.2	364	6	ABU07918 Neisseria
16	210	15.6	323	6	ABU07927 Neisseria
17	209	15.6	391	6	ABU07921 Neisseria
18	209	15.6	393	6	ABU07922 Neisseria
19	209	15.6	398	6	ABU07916 Neisseria
20	209	15.6	400	6	ABU07919 Neisseria
21	206	15.4	325	6	ABU07928 Neisseria
22	204	15.2	405	6	ABU07917 Neisseria
23	204	15.2	405	6	ABU07923 Neisseria
24	204	15.2	407	6	ABU07924 Neisseria
25	204	15.2	407	6	ABU07920 Neisseria

26	204	15.2	645	4	AAE10036	Aae10036 N. mening
27	204	15.2	645	4	AAU27601	Aau27601 Neisseria
28	204	15.2	648	4	AAE10028	Aae10028 N. mening
29	204	15.2	648	4	AAU27579	Aau27579 Neisseria
30	204	15.2	793	4	AAE10016	Aae10016 N. mening
31	204	15.2	793	4	AAU27571	Aau27571 Neisseria
32	204	15.2	806	4	AAE10033	Aae10033 N. mening
33	204	15.2	806	4	AAE10035	Aae10035 N. mening
34	204	15.2	806	4	AAU27598	Aau27598 Neisseria
35	204	15.2	806	4	AAU27600	Aau27600 Neisseria
36	204	15.2	858	4	AAE10019	Aae10019 N. mening
37	204	15.2	858	4	AAU27574	Aau27574 Neisseria
38	204	15.2	1444	4	AAE10037	Aae10037 N. mening
39	204	15.2	1444	4	AAU27602	Aau27602 Neisseria
40	204	15.2	1447	4	AAE10024	Aae10024 N. mening
41	204	15.2	1447	4	AAU27577	Aau27577 Neisseria
42	193	14.4	610	2	AAW68206	Aaw68206 M. catarr
43	193	14.4	624	2	AAW68204	Aaw68204 M. catarr
44	193	14.4	889	2	AAW68208	Aaw68208 M. catarr
45	190.5	14.2	573	2	AAW68202	Aaw68202 M. catarr

ALIGNMENTS

RESULT 1

AAB31713
ID AAB31713 standard; protein; 257 AA.

XX AAB31713;

DT 30-APR-2001 (first entry)

XX Amino acid sequence of the DsrA protein from strain OF406.

DE DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

PD 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX N-PSDB; AAF25270.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.

XX Claim 7; Page 58-59; 80pp; English.

XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in

CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 257 AA;
 Query Match 100.0%; Score 1342; DB 4; Length 257;
 Best Local Similarity 100.0%; Pred. NO. 6.1e-133;
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
 DB 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
 QY 61 KMKPKEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
 DB 61 KMKPKEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITTSKYLLELGTLYLDDSYRMEQNTNINKLS 180
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQNDITTSKYLLELGTLYLDDSYRMEQNTNINKLS 180
 QY 181 KELQTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
 DB 181 KELQTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
 QY 241 NTYNGMSYGASVGVEF 257
 DB 241 NTYNGMSYGASVGVEF 257
 RESULT 2
 AAB31705
 ID AAB31705 standard; protein; 257 AA.
 XX
 AC AAB31705;
 DT 30-APR-2001 (first entry)
 DE Amino acid sequence of the DsrA locus.
 XX
 KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 XX
 OS Haemophilus ducreyi.
 XX
 FN WO200104138-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 07-JUL-2000; 2000WO-US018834.
 XX
 PR 09-JUL-1999; 99US-0143257P.
 XX
 XX (UYNC-) UNIV NORTH CAROLINA.
 PA (ELKI/) ELKINS C.
 XX
 PI Elkins C;
 XX
 DR WPI; 2001-138311/14.
 DR N-PSDB; AAF25262.
 XX
 PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX
 PS Claim 3; Fig 3; 80pp; English.
 PS
 CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing

CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 257 AA;
 Query Match 100.0%; Score 1342; DB 4; Length 257;
 Best Local Similarity 100.0%; Pred. NO. 6.1e-133;
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
 DB 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
 QY 61 KMKPKEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
 DB 61 KMKPKEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITTSKYLLELGTLYLDDSYRMEQNTNINKLS 180
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQNDITTSKYLLELGTLYLDDSYRMEQNTNINKLS 180
 QY 181 KELQTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
 DB 181 KELQTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
 QY 241 NTYNGMSYGASVGVEF 257
 DB 241 NTYNGMSYGASVGVEF 257
 RESULT 3
 AAB31709
 ID AAB31709 standard; protein; 257 AA.
 XX
 AC AAB31709;
 DT 30-APR-2001 (first entry)
 DE Amino acid sequence of the DsrA protein from strain CIP542 (CDC) .
 XX
 KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 XX
 OS Haemophilus ducreyi.
 XX
 FN WO200104138-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 07-JUL-2000; 2000WO-US018834.
 XX
 PR 09-JUL-1999; 99US-0143257P.
 XX
 XX (UYNC-) UNIV NORTH CAROLINA.
 PA (ELKI/) ELKINS C.
 XX
 PI Elkins C;
 XX
 DR WPI; 2001-138311/14.
 DR N-PSDB; AAF25266.
 XX
 PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX
 PS Claim 7; Page 55; 80pp; English.
 PS
 CC The present sequence represents a DsrA 30 kDa protein of Haemophilus

CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 257 AA;

Query Match 100.0%; Score 1342; DB 4; Length 257;
 Best Local Similarity 100.0%; Pred. No. 6.1e-133;
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKIKLVAVVGLACSTTTTAAQQPKFAGVSSLYSYEDYDGKWTWSNEGGFDIKVPGI 60
 DB 1 MKIKLVAVVGLACSTTTTAAQQPKFAGVSSLYSYEDYDGKWTWSNEGGFDIKVPGI 60
 QY 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPDQLGQINRQQLKLN 120
 DB 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPDQLGQINRQQLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTLYLDDSYRMMEQNTNINKLS 180
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTLYLDDSYRMMEQNTNINKLS 180
 QY 181 KELQTGLANOSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
 DB 181 KELQTGLANOSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
 QY 241 NTYNGMSYASGVGEF 257
 DB 241 NTYNGMSYASGVGEF 257

RESULT 4
 AAB31706
 ID AAB31706 standard; protein; 264 AA.

XX AAB31706;
 XX
 XX 30-APR-2001 (first entry)
 XX
 XX Amino acid sequence of the DsrA protein from strain CIPA75.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX N-PSDB; AAF25263.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum

PT resistance to the bacteria used to produce vaccines that induce immune

PT response against the bacteria in subject at risk of developing chancroid.
 XX
 PS Claim 7; Page 53; 80pp; English.

XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 264 AA;

Query Match 96.8%; Score 1299.5; DB 4; Length 264;
 Best Local Similarity 95.8%; Pred. No. 2e-128;
 Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
 QY 1 MKIKLVAVVGLACSTTTTAAQQPKFAGVSSLYSYEDYDGKWTWSNEGGFDIKVPGI 60
 DB 1 MKIKLVAVVGLACSTTTTAAQQPKFAGVSSLYSYEDYDGKWTWSNEGGFDIKVPGI 60
 QY 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPDQLGQINRQQLKLN 120
 DB 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPDQLGQINRQQLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTLYLDDSYRMMEQ-----NT 173
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTLYLDDSYRMMEQNTNINKNT 180
 QY 174 HNINKLSKELQTLGLANOSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFT 233
 DB 181 HNINKLSKELQTLGLANOSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFT 240
 QY 234 AKAGVAFNTYNGMSYASGVGEF 257
 DB 241 AKAGVAFNTYNGMSYASGVGEF 264

RESULT 5
 AAB31707
 ID AAB31707 standard; protein; 264 AA.

XX AAB31707;

XX 30-APR-2001 (first entry)

XX Amino acid sequence of the DsrA protein from strain CIPA77.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX

```

DR WPI; 2001-138311/14.
DR N-PSDB; AAF25264.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 54; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 264 AA;
Query Match 96.8%; Score 1299.5; DB 4; Length 264;
Best Local Similarity 95.8%; Pred. No. 2e-128;
Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
QY 61 KMKPEWISKOATYLELOHYMPTVPLVTYARVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPEWISKOATYLELOHYMPTVPLVTYARVSPSPILLYPMSDDPOLGINRQOLKLN 120
QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQ-----NT 173
Db 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQNTNINKNT 180
QY 174 HNINKLSKELQTLGANSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGSRTDRFT 233
Db 181 HNINKLSKELQTLGANSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGSRTDRFT 240
QY 234 AKAGVAFNTYNGGMSYGASVGYEF 257
Db 241 AKAGVAFNTYNGGMSYGASVGYEF 264
RESULT 6
AAB31711
ID AAB31711 standard; protein; 271 AA.
XX
XX AAB31711;
XX
XX 30-APR-2001 (first entry)
XX
XX Amino acid sequence of the DsrA protein from strain V-1157.
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
XX Haemophilus ducreyi.
XX
XX WO200104138-A1.
XX
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000WO-US018834.
XX
XX 09-JUL-1999; 99US-0143257P.
XX
XX

```

```

PA (UYNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
XX WPI; 2001-138311/14.
DR N-PSDB; AAF25268.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 57; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 271 AA;
Query Match 96.1%; Score 1289; DB 4; Length 271;
Best Local Similarity 93.0%; Pred. No. 2.6e-127;
Matches 252; Conservative 0; Mismatches 5; Indels 14; Gaps 1;
QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
QY 61 KMKPEWISKOATYLELOHYMPTVPLVTYARVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPEWISKOATYLELOHYMPTVPLVTYARVSPSPILLYPMSDDPOLGINRQOLKLN 120
QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQ----- 171
Db 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQNTNINKNT 180
QY 172 -----NTHNINKLSKELQTLGANSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGS 226
Db 181 HNINKNTHNINKLSKELQTLGANSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGS 240
QY 227 RITDRFTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 241 RITDRFTAKAGVAFNTYNGGMSYGASVGYEF 271
RESULT 7
AAB31712
ID AAB31712 standard; protein; 273 AA.
XX
XX AAB31712;
XX
XX 30-APR-2001 (first entry)
XX
XX Amino acid sequence of the DsrA protein from strain M90-02.
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
XX Haemophilus ducreyi.
XX
XX WO200104138-A1.
XX
XX 18-JAN-2001.
XX

```

```

XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX XX (UYNC-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX XX
XX PI Elkins C;
XX XX
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25269.
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX PT resistance to the bacteria used to produce vaccines that induce immune
XX PT response against the bacteria in subject at risk of developing chancroid.
XX PS Claim 7; Page 58; 80pp; English.
XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX CC polypeptide is used to produce vaccine compositions, which are useful for
XX CC inducing a protective immune response in a subject at risk of developing
XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX CC fragments or peptides can be used for screening libraries of compounds in
XX CC a variety of drug screening techniques. The proteins and peptides may be
XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
XX CC The nucleic acids are useful for the preparation of DsrA proteins
XX SQ Sequence 273 AA;

Query Match 95.6%; Score 1283; DB 4; Length 273;
Best Local Similarity 92.7%; Pred. No. 1.1e-126;
Matches 253; Conservative 0; Mismatches 4; Indels 16; Gaps 2;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKAGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKAGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTVYAGVSPS--PILLYPMSDDPDLGINRQOLK 118
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTVSAPDVSPSSIILLYPMSDDPDLGINRQOLK 120
Qy 119 LNLVSYFNDLRHDFKLVLDARISKNQNDITISKYLLLELGTYLDDSVRMWQ----- 171
Db 121 LNLVSYFNDLRHDFKLVLDARISKNQNDITISKYLLLELGTYLDDGSYRMEQNTNINK 180
Qy 172 -----NTHNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGV 224
Db 181 NTHNINKNTNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGV 240
Qy 225 GSRITDRFTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 241 GSRITDRFTAKAGVAFNTYNGGMSYGASVGYEF 273

RESULT 8
AAB31708
ID AAB31708 standard; protein; 242 AA.
XX AC
XX AC AAB31708;
XX DT 30-APR-2001 (first entry)
XX XX Amino acid sequence of the DsrA protein from strain CIP542 (Can).
XX DE DsrA protein; outer membrane protein; serum resistance; vaccine;
XX KW chancroid disease; genital ulcer disease.
XX XX

```

```

OS Haemophilus ducreyi.
XX WO200104138-A1.
XX PD 18-JAN-2001.
XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX XX (UYNC-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX XX
XX PI Elkins C;
XX XX
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25265.
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX PT resistance to the bacteria used to produce vaccines that induce immune
XX PT response against the bacteria in subject at risk of developing chancroid.
XX PS Claim 7; Page 55; 80pp; English.
XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX CC polypeptide is used to produce vaccine compositions, which are useful for
XX CC inducing a protective immune response in a subject at risk of developing
XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX CC fragments or peptides can be used for screening libraries of compounds in
XX CC a variety of drug screening techniques. The proteins and peptides may be
XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
XX CC The nucleic acids are useful for the preparation of DsrA proteins
XX SQ Sequence 242 AA;

Query Match 93.8%; Score 1259; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.3e-124;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKAGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKAGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTVYAGVSPS--PILLYPMSDDPDLGINRQOLK 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTVYAGVSPS--PILLYPMSDDPDLGINRQOLK 120
Qy 121 LYSYFNDLRHDFKLVLDARISKNQNDITISKYLLLELGTYLDDSVRMEQNTNINKLS 180
Db 121 LYSYFNDLRHDFKLVLDARISKNQNDITISKYLLLELGTYLDDSVRMEQNTNINKLS 180
Qy 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
Db 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
Qy 241 NT 242
Db 241 NT 242

RESULT 9
AAB31710
ID AAB31710 standard; protein; 263 AA.
XX AC
XX AC AAB31710;
XX DT 30-APR-2001 (first entry)
XX XX

```



```

XX 23-MAY-2003 (first entry)
DT XX
DE XX
DE Neisserial adhesin A (NadA) allele 1.
XX
XX Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;
KW Neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW Neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW systemic immunity; mucosal immunity; allele; strain 95330; chimeric.
XX
XX Neisseria meningitidis.
OS Synthetic.
OS
XX WO2003010194-A2.
XX
XX 06-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-IB003396.
XX
XX 27-JUL-2001; 2001GB-00018401.
XX
XX 06-SEP-2001; 2001GB-00021591.
XX
XX 14-MAY-2002; 2002GB-00011025.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Arico M, Comanducci M;
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding Neisserial
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EY-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC Neisserial adhesin A (NadA) allele 1/2 (first ATG start) chimera (strain
CC 95330)
XX
XX Sequence 357 AA;
XX
Query Match 16.2%; Score 217; DB 6; Length 357;
Best Local Similarity 39.4%; Pred. No. 8.7e-14;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;
QY 127 DLRHDFKLVLDARISKKNQIDITISKYLLGLTYLDDSYRMMEQNTNINKLSKELQTG 186
DQ 249 DIKADIATNKAD--IAKNSARIDSLDK-----NVANLRKETRQG 285
QY 187 LANQSAISMLVQPNVGKTSVSAAGGYRDKTALAIGVGSRTIDRFTAKGVAFTNYG- 245
DQ 286 LAEQAALSGLFQPNVGRFNTAAVGGYKSESVAIGTGFTFTENFAKAGVAVGTSSGS 345
QY 246 GMSYGASVGYEF 257
DQ 346 SAAYHVGVNVEW 357
RESULT 12
ABU07915
ID ABU07915 standard; protein; 362 AA.
XX
XX ABU07915;
XX

```

```

DT 23-MAY-2003 (first entry)
DE XX
DE Neisserial adhesin A (NadA) allele 1.
XX
XX Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;
KW Neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW Neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW systemic immunity; mucosal immunity; allele.
XX
XX Neisseria meningitidis.
OS
XX WO2003010194-A2.
XX
XX 06-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-IB003396.
XX
XX 27-JUL-2001; 2001GB-00018401.
XX
XX 06-SEP-2001; 2001GB-00021591.
XX
XX 14-MAY-2002; 2002GB-00011025.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Arico M, Comanducci M;
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 75; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding Neisserial
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EY-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of
CC Neisserial adhesin A (NadA) allele 1
XX
XX Sequence 362 AA;
XX
Query Match 16.2%; Score 217; DB 6; Length 362;
Best Local Similarity 39.4%; Pred. No. 8.9e-14;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;
QY 127 DLRHDFKLVLDARISKKNQIDITISKYLLGLTYLDDSYRMMEQNTNINKLSKELQTG 186
DQ 254 DIKADIATNKAD--IAKNSARIDSLDK-----NVANLRKETRQG 290
QY 187 LANQSAISMLVQPNVGKTSVSAAGGYRDKTALAIGVGSRTIDRFTAKGVAFTNYG- 245
DQ 291 LAEQAALSGLFQPNVGRFNTAAVGGYKSESVAIGTGFTFTENFAKAGVAVGTSSGS 350
QY 246 GMSYGASVGYEF 257
DQ 351 SAAYHVGVNVEW 362
RESULT 13
AAY75736
ID AAY75736 standard; protein; 364 AA.
XX
XX AAY75736;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 961 protein sequence SEQ ID NO:2944.

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:12 ; Search time 22 Seconds
(without alignments)
603.085 Million cell updates/sec

Title: US-10-030-529a-2

Perfect score: 1342

Sequence: 1 MKIKLVAVVGLACSTITM.....VAFTNYNGMSYGASGVYEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	14.4	610	4	US-09-336-447A-11
2	193	14.4	624	4	US-09-336-447A-7
3	193	14.4	889	4	US-09-336-447A-15
4	192	14.3	867	4	US-09-540-236-2676
5	190.5	14.2	573	4	US-09-336-447A-3
6	125.5	9.4	568	4	US-09-543-681A-6966
7	115	8.6	2042	4	US-09-077-098A-6
8	114	8.5	2039	4	US-09-077-098A-7
9	110	8.2	873	4	US-09-336-447A-13
10	108	8.0	892	4	US-09-336-447A-5
11	107	8.0	831	4	US-09-336-447A-1
12	107	8.0	878	4	US-09-540-236-3401
13	107	8.0	941	4	US-09-336-447A-9
14	104	7.7	1002	4	US-09-268-347-24
15	104	7.7	1004	4	US-09-268-347-30
16	103.5	7.7	616	4	US-09-268-347-38
17	103	7.7	1104	4	US-09-268-347-28
18	103	7.7	1104	4	US-09-268-347-34
19	102	7.6	679	3	US-08-913-942-15
20	102	7.6	679	4	US-09-268-347-26
21	102	7.6	1094	4	US-09-268-347-32
22	100.5	7.5	1098	1	US-08-409-995-2
23	100.5	7.5	1098	3	US-08-685-467-2
24	100.5	7.5	1098	3	US-09-377-155-32
25	100.5	7.5	1098	3	US-08-913-942-2
26	100.5	7.5	1098	4	US-09-669-974-32
27	100.5	7.5	1098	4	US-09-268-347-44

28 100.5 7.5 1098 4 US-09-797-862-32 Sequence 32, Appl
29 100.5 7.5 2353 3 US-09-377-355-33 Sequence 33, Appl
30 100.5 7.5 2353 3 US-08-913-942-4 Sequence 4, Appl
31 100.5 7.5 2353 4 US-09-669-974-33 Sequence 33, Appl
32 100.5 7.5 2353 4 US-09-797-862-33 Sequence 33, Appl
33 100.5 7.5 2411 4 US-09-268-347-36 Sequence 36, Appl
34 99 7.4 1690 4 US-09-595-684B-39 Sequence 39, Appl
35 99 7.4 2354 4 US-09-268-347-47 Sequence 47, Appl
36 93 6.9 1117 2 US-08-843-530B-33 Sequence 33, Appl
37 92 6.9 447 4 US-09-543-681A-6231 Sequence 6231, Appl
38 90 6.7 512 4 US-09-489-039A-12836 Sequence 12836, A
39 88 6.6 776 4 US-10-181-660-5 Sequence 5, Appl
40 88 6.6 776 4 US-10-181-660-13 Sequence 13, Appl
41 88 6.6 797 4 US-09-994-192-4 Sequence 4, Appl
42 88 6.6 797 4 US-10-181-660-3 Sequence 3, Appl
43 86.5 6.6 797 4 US-10-181-660-11 Sequence 11, Appl
44 1226 1 US-08-280-443-2 Sequence 2, Appl
45 86.5 6.4 1226 1 US-08-457-459-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-336-447A-11
; Sequence 11, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-11

Query Match 14.4%; Score 193; DB 4; Length 610;
Best Local Similarity 37.9%; Pred. No. 4.2e-12;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;
Qy 126 NDLRHDFKLVLDARISKKNQNTIDTSKYLLELGTYLDD-SYRMEQNTHT-----NINK 178
Db 471 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVKDFGRVLTALDTKVNAPDGRITA 530
Qy 179 LSXELQTLANQSNLSMLVQPNVGKTSVSNAGGYRDKTALAIGVCSRTIDRTAKGV 238
Db 531 LSKVNGMAQAALSGLFQFYSVGKFNATAALGGYSGSAVATGAGYRVPNPLAFKAGA 590
Qy 239 AFNTY-NGGMSYGASGVYEF 257
Db 591 AINTSGNKGSYNIGVNYEF 610

RESULT 2

US-09-336-447A-7
; Sequence 7, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.

; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP21 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-7

Query Match 14.4%; Score 193; DB 4; Length 624;
Best Local Similarity 37.9%; Pred. No. 4.3e-12;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;
QY 126 NDLRHFPLKVLDAIRISKKNQIDITISKYLLLELGTLYDD-SYRMEQNTN-178
Db 485 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVDFGDRVTALDTKVNAPDGRITA 544
QY 179 LSKELQTLGANSALSMVQPNVGKTSVSAAGGYRDKTALAIGVGSRIIDRFTAKGV 238
Db 545 LDSKVENGMAQAALSGLFQPSYVGKFNATAALGGYGSKSAVAIGAGYRVPNPLAFKAGA 604
QY 239 AFNTY-NGGMSYGASGVYEF 257
Db 605 AINTSGNKKGSYNGVNYEF 624

RESULT 3
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP21 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 14.4%; Score 193; DB 4; Length 889;
Best Local Similarity 37.9%; Pred. No. 7.5e-12;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;
QY 126 NDLRHFPLKVLDAIRISKKNQIDITISKYLLLELGTLYDD-SYRMEQNTN-178
Db 750 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVDFGDRVTALDTKVNAPDGRITA 809
QY 179 LSKELQTLGANSALSMVQPNVGKTSVSAAGGYRDKTALAIGVGSRIIDRFTAKGV 238
Db 810 LDSKVENGMAQAALSGLFQPSYVGKFNATAALGGYGSKSAVAIGAGYRVPNPLAFKAGA 869
QY 239 AFNTY-NGGMSYGASGVYEF 257
Db 870 AINTSGNKKGSYNGVNYEF 889

RESULT 4
US-09-540-236-2676
; Sequence 2676, Application US/09540236

; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2676
; LENGTH: 867
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2676

Query Match 14.3%; Score 192; DB 4; Length 867;
Best Local Similarity 37.9%; Pred. No. 9.3e-12;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;
QY 126 NDLRHFPLKVLDAIRISKKNQIDITISKYLLLELGTLYDD-SYRMEQNTN-178
Db 728 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVDFGDRVTALDTKVNAPDGRITA 787
QY 179 LSKELQTLGANSALSMVQPNVGKTSVSAAGGYRDKTALAIGVGSRIIDRFTAKGV 238
Db 788 LDSKVENGMAQAALSGLFQPSYVGKFNATAALGGYGSKSAVAIGAGYRVPNPLAFKAGA 847
QY 239 AFNTY-NGGMSYGASGVYEF 257
Db 848 AINTSGNKKGSYNGVNYEF 867

RESULT 5
US-09-336-447A-3
; Sequence 3, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP21 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-3

Query Match 14.2%; Score 190.5; DB 4; Length 573;
Best Local Similarity 35.4%; Pred. No. 7.1e-12;
Matches 52; Conservative 18; Mismatches 62; Indels 15; Gaps 3;
QY 126 NDLRHFPLKVLDAIRISKKNQIDITISKYLLLELGTLYDD-SYRMEQNTN-174
Db 427 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVDFGDRVTALDTKVNALDTKVNA 486
QY 175 ---NINKLSKELOTLGANSALSMVQPNVGKTSVSAAGGYRDKTALAIGVGSRIIDR 231
Db 487 FDGRITALDSKVENGMAQAALSGLFQPSYVGKFNATAALGGYGSKSAVAIGAGYRVPN 546
QY 232 FTAKAGVAFNTY-NGGMSYGASGVYEF 257
Db 547 LAFKAGVAFNTY-NGGMSYGASGVYEF 573

RESULT 6

US-09-543-681A-6966
; Sequence 6966, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6966
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-6966

Query Match 9.4%; Score 125.5; DB 4; Length 568;
Best Local Similarity 26.18; Pred. No. 8.7e-05;
Matches 42; Conservative 21; Mismatches 49; Indels 49; Gaps 5;

Qy 137 LDARISKKNQIDT-----ISKYLL-----ELGTLYLDDSYRM 168
Db 404 VESDVKNKEDIQTNRENINQVKQTADKNKIYVDKYLANNFDVNEQSKSLVGLSLYAG 463

Qy 169 MEQNTNIN-----KLSKELOTGLANOSALSMVLVOP-NGVKTTSVSA 209
Db 464 NKSNTDINAINKDLSHFQNETNRRFYKVKRANQGIASVAAMSNL--PFNDAAFTSTAM 521

Qy 210 AVGGYRDKTALAIGVGRITDRFTAKAGVAFNTYNGGMSG 250
Db 522 GIGNYNATAFAMQYRINENVKVKASTAWNDANNWVSAG 562

RESULT 7
US-09-077-098A-6
; Sequence 6, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098A
; FILING DATE: 19-May-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

US-09-077-098A-6
; Sequence 6, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098A
; FILING DATE: 19-May-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-077-098A-6

Query Match 8.6%; Score 115; DB 4; Length 2042;
Best Local Similarity 29.8%; Pred. No. 0.009;
Matches 31; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 157 ELGTLYLDDSYRMMEQNTHNINKLSKELQTLGLANOSALSMVLVQPNVGKTSVSAAVGGYRD 216
Db 1942 QLNTVIDNVQNNFNQVRIGDLTRESRAGIAGAMATASLQNVALPKGTTISVGTATFKG 2001

Qy 217 KTLATGVGRITDRFTAKAGV---APTNYNGGMSGYGVYEF 257
Db 2002 ENAVAIGM-SRLSD--NGKVGIRLSGMSGTNGDKGAAMSVGF 2042

RESULT 8
US-09-077-098A-7
; Sequence 7, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098A
; FILING DATE: 19-May-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

US-09-077-098A-6
; Sequence 6, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098A
; FILING DATE: 19-May-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

US-09-077-098A-6
; Sequence 6, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098A
; FILING DATE: 19-May-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

```
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-077-098A-7

Query Match      8.5%; Score 114; DB 4; Length 2039;
Best Local Similarity 29.8%; Pred. No. 0.012;
Matches 31; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 157 ELGTYLDDSYRMMEQNTHNINKLSKELQTLGLANQSALSMVOPNGVGTSTVSAAVGGYRD 216
Db 1939 QLVTVIDNVQNNQVNRIGDLTRESRAGIAGAVATASLQNVALPGKTTISVGTATPKG 1998

Qy 217 KTLAIGVGSRTIDRTAKAGV---AFNTYNGMSYGSVGYEF 257
Db 1999 ENAVAIGM-SRLSD--NGKVGIRLSGMSSTSGDKGAANSVGFTF 2039

RESULT 9
US-09-336-447A-13
; Sequence 13, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-13

Query Match      8.2%; Score 110; DB 4; Length 873;
Best Local Similarity 27.6%; Pred. No. 0.0083;
Matches 42; Conservative 30; Mismatches 52; Indels 28; Gaps 8;

Qy 110 LGINROOLKLNLYSFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTYLDDSYRMM 169
Db 745 IATNKOELILQ-----ILQHD-RLNRINETNNRQDKID-----QLG-----YALK 781

Qy 170 EQNTHNINKLSK-ELQTL--GLANQSALSMVOPNGVGTSTVSAAVGGYRDKTALAIG-VG 225
Db 782 EQQHFNRRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 841

Qy 226 SRTIDRTAKAGVAFNTYNGMSYGSVGYEF 257
Db 842 LSDTGKSTYKIGLSWSD-AGGLSGVGGSYRW 872

RESULT 10
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match      8.0%; Score 108; DB 4; Length 892;
Best Local Similarity 27.6%; Pred. No. 0.014;
Matches 42; Conservative 29; Mismatches 53; Indels 28; Gaps 8;

Qy 110 LGINROOLKLNLYSFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTYLDDSYRMM 169
Db 764 IATNKOELILQ-----ND-----RLNQINETNNRQDKID-----QLG-----YALK 800

Qy 170 EQNTHNINKLSK-ELQTL--GLANQSALSMVOPNGVGTSTVSAAVGGYRDKTALAIG-VG 225
Db 801 EQQHFNRRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 860

Qy 226 SRTIDRTAKAGVAFNTYNGMSYGSVGYEF 257
Db 861 LSDTGKSTYKIGLSWSD-AGGLSGVGGSYRW 891

RESULT 11
US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Query Match      8.0%; Score 107; DB 4; Length 831;
Best Local Similarity 27.6%; Pred. No. 0.016;
Matches 42; Conservative 29; Mismatches 53; Indels 28; Gaps 8;

Qy 110 LGINROOLKLNLYSFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTYLDDSYRMM 169
Db 703 IATNKOELILQ-----ND-----RLNRINETNNRQDKID-----QLG-----YALK 739

Qy 170 EQNTHNINKLSK-ELQTL--GLANQSALSMVOPNGVGTSTVSAAVGGYRDKTALAIG-VG 225
Db 740 EQQHFNRRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 799

Qy 226 SRTIDRTAKAGVAFNTYNGMSYGSVGYEF 257
Db 800 LSDTGKSTYKIGLSWSD-AGGLSGVGGSYRW 830

RESULT 12
US-09-540-236-3401
; Sequence 3401, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
```

```

; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3401
; LENGTH: 878
; TYPE: prt
; ORGANISM: M.catarrhalis
US-09-540-236-3401

```

Query Match	8.0%;	Score 107;	DB 4;	Length 878;
Best Local Similarity	27.6%;	Pred. No. 0.018;		
Matches	42;	Conservative 29;	Mismatches 53;	Indels 28; Gaps 8;

Qy	110	LGINRQQLKLNLYSYFNDLRHDFKLKVLDA	RI	SK	KN	K	N	I	D	T	S	K	V	L	E	L	G	T	L	D	S	R	M	169
Db	750	IATNKQLLIQ-----ND-----	KL	N	R	I	N	E	T	N	N	Q	D	K	I	D	-----	OLG-----	Y	A	L	K	786	
Qy	170	EQMTNINIKLSK-ELQT--GLANOSALSM	L	V	O	P	N	G	V	G	K	T	S	V	S	A	A	V	G	V	R	K	T	225
Db	787	EQCHFNRIISAVERTQTAGGIANAIA	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	
Qy	226	SRITDRFTKAGVAFNTYNGM	S	Y	C	A	S	V	G	Y	E	F	257											
Db	847	LSPTGKSTYKIGLSWSD-AGLSGGVGG	S	Y	R	W	877																	

```

RESULT 13
US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336-447A-9
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

```

	Query Match	8.0%;	Score 107;	DB 4;	Length 941;
	Best Local Similarity	25.9%;	Pred. No. 0.02;		
	Matches	41;	Conservative	36;	Mismatches 63; Indels 18; Gaps 8;
QY	113	NRQOLKLNLYSYENDLRHDFKLVLDARISQKNQNI-----DTISKYLLLELGTLYLDD-----	164		
Db	788	NKTAIEQNIN---RTVANGPEIEIKRKAGIATNKOELIQLNDRLNQ--INETNNHQDQKIDQ	843		
QY	165	-SYRMMEQNTHTNINKLSK-ELQT--GLANGSALSMLVQPNQGVGKTSYSAAVGGVYDKTAL	220		
Db	844	LGVALKEQGGHFNNRIISAVERTAGGTANAIAIATLPSPAGEHHVLFSGSYHNGQAAV	903		
QY	221	AIG-VGSRITDRFTAKAGVAFNTYNGMSYGCASVGYEF	257		
Db	904	SILGAAGHSIDTKSTFKYKIGLSWSD--AGGLSGGVGGVSFEW	940		

RESULT 14
US-09-268-347-24
; Sequence 24, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEM
; FILE REFERENCE: 1038-860

```

: CURRENT APPLICATION NUMBER: US/09/268,347
:
: CURRENT FILING DATE: 1999-03-16
:
: NUMBER OF SEQ ID NOS: 54
:
: SOFTWARE: Patentin Ver. 2.0
:
: SEQ ID NO 24
:
: LENGTH: 1002
:
: TYPE: PRT
:
: ORGANISM: Haemophilus influenzae
: US-09-268-347-24

```

[illegible]

```

RESULT 15
US-09-268-347-30
; Sequence 30, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-30

```

Query Match	7.7%	Score 104;	DB 4;	Length 1004;
Best Local Similarity	36.1%;	Pred. No. 0.047;		
Matches	30;	Conservative	15;	Mismatches 36;
				Indels 2;
				Gaps 2;
Qy	176	INKLSKELOTGLANQSA	LSMLVGVNGV	GVKTSVSAAVGGYRDKTALAIGVSRITDRFTAK
		:::::	:::::	:::::
Db	919	VNVKGRADAGTASALA	ASQLPQASMPGKSMVSI	AGSSYQQSGSLAIGV-SRISDNGKVI
		:::::	:::::	:::::
Qy	236	AGVAENTYNGGMS-	YGASVGYEF	257
		:::::	:::::	:::::
Db	978	IRLSGTTNSOGKTVG	AAAGVGYOW	1000
		:::::	:::::	:::::

Search completed: May 13, 2004, 08:18:02
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:16:37 ; Search time 48 Seconds
(without alignments)
1489.858 Million cell updates/sec

Title: US-10-030-529A-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFTNYGMSYGASGVYEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	14.4	610	10	US-09-952-267-11
2	193	14.4	624	10	US-09-952-267-7
3	193	14.4	889	10	US-09-952-267-15
4	190.5	14.2	573	10	US-09-952-267-3
5	159.5	11.9	2712	12	US-10-282-122A-67070
6	115	8.6	2042	14	US-10-192-584-6
7	114	8.5	2039	14	US-10-192-584-7
8	110	8.2	873	10	US-09-952-267-13
9	108	8.0	852	12	US-10-282-122A-62892
10	108	8.0	892	10	US-09-952-267-5
11	107	8.0	831	10	US-09-952-267-1
12	107	8.0	941	10	US-09-952-267-9
13	100.5	7.5	1098	9	US-09-797-862-32
14	100.5	7.5	2353	9	US-09-797-862-33
15	97	7.2	496	9	US-09-738-626-5483

16	96.5	7.2	795	12	US-10-282-122A-63527	Sequence 63527, A
17	90.5	6.7	723	15	US-10-369-493-10942	Sequence 10942, A
18	90.5	6.7	819	12	US-10-282-122A-54896	Sequence 54896, A
19	90.5	6.7	3073	12	US-10-282-122A-49147	Sequence 49147, A
20	90	6.7	478	9	US-09-815-242-11674	Sequence 11674, A
21	90	6.7	512	12	US-10-282-122A-59775	Sequence 59775, A
22	90	6.7	1055	14	US-10-032-585-7652	Sequence 7652, Ap
23	89.5	6.7	1074	12	US-10-282-122A-50616	Sequence 50616, A
24	88	6.6	335	12	US-10-282-122A-60846	Sequence 60846, A
25	88	6.6	797	15	US-10-320-800-6	Sequence 6, Appl1
26	88	6.6	1246	12	US-10-282-122A-49773	Sequence 49773, A
27	87	6.5	827	12	US-10-282-122A-60889	Sequence 60889, A
28	86.5	6.4	400	15	US-10-369-493-10982	Sequence 10982, A
29	86.5	6.4	1107	12	US-10-282-122A-75398	Sequence 75398, A
30	86.5	6.4	1225	14	US-10-233-553-20	Sequence 20, Appl
31	86	6.4	473	12	US-10-282-122A-76706	Sequence 76706, A
32	86	6.4	652	12	US-10-282-122A-49839	Sequence 49839, A
33	85.5	6.4	1461	12	US-10-282-122A-75339	Sequence 75339, A
34	85	6.3	278	12	US-10-424-599-264683	Sequence 264683, A
35	85	6.3	633	12	US-10-282-122A-52615	Sequence 52615, A
36	84.5	6.3	356	9	US-09-771-382-37	Sequence 37, Appl
37	84.5	6.3	382	9	US-09-771-382-38	Sequence 38, Appl
38	84.5	6.3	407	9	US-09-771-382-25	Sequence 25, Appl
39	84.5	6.3	433	9	US-09-771-382-26	Sequence 26, Appl
40	84.5	6.3	461	9	US-09-771-382-35	Sequence 35, Appl
41	84.5	6.3	502	9	US-09-771-382-27	Sequence 27, Appl
42	84.5	6.3	512	9	US-09-771-382-23	Sequence 23, Appl
43	84.5	6.3	540	9	US-09-771-382-33	Sequence 33, Appl
44	84.5	6.3	591	9	US-09-797-862-11	Sequence 11, Appl
45	84.5	6.3	591	9	US-09-797-862-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1

US-09-952-267-11
; Sequence 11, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-11

Query Match 14.4%; Score 193; DB 10; Length 610;

Best Local Similarity 37.9%; Pred. No. 1.5e-10;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;

Qy	126	NDLRHPFKLVLDARISKNQNTDTTSKYLLEGTGLDD-SYRMEQNTH-----NINK 178
Db	471	NKASDTKFAATADATKNGNAITKNKASITDGLTKVDGFGRTALDTKNAPDGRITA 530
Qy	179	LSKELQTGLANQSAISMLVQPNVGKTSVSAAGVGRDKTALAIQVGSRTIDRETKAGV 238
Db	531	LDSKVENGAQAALSGLFQPSYGVGNATAALGGYGSKAVALGAGYRVNPNLAFKAGA 590
Qy	239	AFNTY-NGMSYGASGVYEF 257

Quest
Best

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 20 Seconds
(without alignments)
1236.061 Million cell updates/sec

Title: US-10-030-529a-2
Perfect score: 1342
Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFTNYGMSYGASGVYEF 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	217	16.2	364	2	adhesin/invasin, p
2	186.5	13.9	455	2	yopA protein - Yer
3	186	13.9	422	2	yopA protein - Yer
4	182	13.6	434	2	invasin precursor
5	132	9.8	155	2	immunoglobulin-bin
6	119.5	8.9	658	2	probable surface p
7	112.5	8.4	364	2	probable exported
8	109.5	8.2	338	2	adhesin/invasin-li
9	109.5	8.2	338	2	adhesin/invasin-li
10	102.5	7.6	622	2	probable exported
11	99.5	7.4	831	2	hypothetical prote
12	97.5	7.3	1695	2	hypothetical prote
13	97	7.2	466	2	hypothetical prote
14	96.5	7.2	795	1	endopeptidase Ia (
15	95	7.1	466	2	hypothetical prote
16	93.5	7.0	163	2	ORF MSV219 hypote
17	92.5	6.9	588	2	hypothetical prote
18	91	6.8	699	2	conserved hypote
19	91	6.8	765	2	hypothetical prote
20	91	6.8	2715	2	eyelid - fruit fly
21	90.5	6.7	340	2	probable periplasm
22	90.5	6.7	819	1	endopeptidase Ia (
23	90.5	6.7	819	2	Lon ATP-dependent
24	90	6.7	1328	1	retrovirus-related
25	89.5	6.7	956	2	excinuclease ABC (
26	89	6.6	525	2	conserved hypote
27	89	6.6	1588	2	probable adhesin Z
28	89	6.6	1588	2	probable adhesin E
29	88.5	6.6	498	2	flagellar hook-ass

30	88	6.6	335	2	AF11125	phosphate transpor
31	88	6.6	530	2	T19365	hypothetical prote
32	88	6.6	797	2	D82000	outer membrane pro
33	88	6.6	797	2	G81228	outer membrane pro
34	88	6.6	4543	1	A53102	alpha-2-macroglobu
35	87.5	6.5	1032	2	B86224	hypothetical prote
36	87	6.5	335	2	AE1486	phosphate transpor
37	87	6.5	658	2	T24851	hypothetical prote
38	87	6.5	806	1	E70131	endopeptidase Ia (
39	87	6.5	827	2	AD1311	penicillin-binding
40	87	6.5	1190	2	A82615	surface protein Xf
41	86.5	6.4	418	2	A53120	intracellular coag
42	86.5	6.4	524	2	T07988	ATP-dependent clp
43	86.5	6.4	1107	2	AC0976	probable autotrans
44	86.5	6.4	1226	1	S65593	adenosine deaminas
45	86	6.4	371	2	C90294	hypothetical prote

ALIGNMENTS

RESULT 1

A:81019
adhesin/invasin, probable NMB1994 [imported] - Neisseria meningitidis (strain MC58 serog:
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81019
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: A81019
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <YET>
A:Cross-references: GB:AE002548; GB:AE002098; NID:G7272749; PIDN:AAF42321.1; PID:G7272725;
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1994

Query Match	16.2%	Score 217;	DB 2;	Length 364;
Best Local Similarity	39.4%	Pred. No. 5.3e-10;		
Matches	52;	Conservative	19;	Mismatches 37; Indels 24; Gaps 3;
Qy	127	DLRHDFKLKVLDAISKQKQNTIDTISKYLELGTYLDDSYRMMEQNTNINKLSKELQTG	186	
Db	256	DIKADTATNKAD--IAKNSARIDSLDK-----NVANLRKETRQG	292	
Qy	187	LANQSALSMVLQPNQGVKTSVSAAGGYRDKTALAIGVGSRIITDRFTAKAGVAFNTVNG-	245	
Db	293	LAEQAALSGLFQPNVGRFNVTAAVGVKSESVAIGTGFRFTENFAAKGAVAGVTSSGS	352	
Qy	246	GMSYGASGVYEF	257	
Db	353	SAAYHGVNVEW	364	

RESULT 2

S04912
yopA protein - Yersinia enterocolitica plasmid pYV6471/76
C:Species: Yersinia enterocolitica
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S04912
R:Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A:Reference number: S04910; MUID:89343638; PMID:2761389
A:Accession: S04912
A:Status: preliminary
A:Molecule type: DNA

A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0110
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-658 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:gl5978974; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0902

Query Match 8.9%; Score 119.5; DB 2; Length 658;
Best Local Similarity 27.4%; Pred. No. 0.1;
Matches 31; Conservative 25; Mismatches 50; Indels 7; Gaps 4;

QY 147 NIDTISKYLLLELGYLDDSYRMMEONTHNINKLSKELQTLGLANQSALSMVQPNVGVKTS 206
DB 551 NFDOLKISINQWYATNQRVSELKQ---DIRKQNSVLASIASMSVASITQPTYSSTM 607

QY 207 VSAAGVGRDKTALAIGVSRITD--RFTAKAGVAFNTYNGMSYGASVGYEF 257
DB 608 TTIGAASYRGQSALSLGVSS-ISDSGRWVKLQASSNT-QGDFGIGVGVGVQW 658

RESULT 7
AB0169
probable exported protein YPO1387 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AB0169
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0169
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90216.1; PID:gl5979436; GSPDB:GN00175
C;Genetics:
A;Gene: YPO1387

Query Match 8.4%; Score 112.5; DB 2; Length 364;
Best Local Similarity 26.8%; Pred. No. 0.16;
Matches 40; Conservative 25; Mismatches 69; Indels 15; Gaps 6;

QY 111 GINRQQLKALYSYFNDLRHDFKLK--VLDARISKKNKONITDTISKYLLLELGYLDDSYRM 168
DB 226 GKSKETLN-NTYDIV-----DSKVSSIVYDVNSYTDKTVNTAPETSLSDAKSYVDKYY-- 277

QY 169 MEQNTNINKLSKELQTLGLANQSALSMVQPNVGVKTSVSAAGVGRDKTALAIGVGSRI 228
DB 278 -NQLSDKVNKNFNKNKNAGISGAMAMSGIPQFGYK-SFGMAIGAYRGQSALAVGGDWNI 335

QY 229 TDRFTAKAGVAFNTYNGMSYGASVGYEF 257
DB 336 NHIKTIIRVNSADTEGG---VGVAAGFAF 361

RESULT 8
D90697
adhesin/invasin-like protein [imported] - Escherichia coli (strain O157:H7, substrain R)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90697
R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome reference number: A99629; MUID:21156231; PMID:11258796
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90697
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-338 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33971.1; PID:gl3360006; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: ECs0548

Query Match 8.2%; Score 109.5; DB 2; Length 338;
Best Local Similarity 20.1%; Pred. No. 0.26;
Matches 56; Conservative 32; Mismatches 92; Indels 99; Gaps 9;

QY 73 TYLELQHYMPYTPVLVTYAPGVSPSP-----ILLYPMSDDP 108
DB 65 TYLTEHHYIPSETPTTQTPTVQTPDAGQKTAATGDVQTTARYQSMINAROSAVTDAQ 124

QY 109 QLGINRQQLK-----NLYSY-----FNDLR 129
DB 125 OTQITTEQQAIVATQKTLAATGDTQNTAHYQEMINARLAQNEANQRTATEQGGKNALT 184

QY 130 HDFKLKVLDAISKNKQ-----NIDTISKYLLLELGYLDDSYRMMEQNTHNI 176
DB 185 TDVAVQQONERTQYDKMQSLAQESAQAHEQIDLSQDVQTTQHQLTNTQKRVADNSQOI 244

QY 177 NKLS-----KELOTGLANQSALSMVQPNVGVKTS-----VSAAGVGRDKT 218
DB 245 NTLNNHFFSLKNEVDNRRKEANAGTA--SAIAIASQPQ--VKTGDMVMVMSAGAGTFNGES 300

QY 219 ALAIGVGSRTITDFTAKAGVAFNTYNGMSYGASVGYEF 257
DB 301 AVSVGTGFNAGTHVLKAGISADT-QSDFGAGVGVGYSF 338

RESULT 9
G85547
adhesin/invasin-like protein [similarity] - Escherichia coli (strain O157:H7, substrain R)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C;Accession: G85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-338 <STO>
A;Cross-references: GB:AE005174; MID:gl2513376; PIDN:AAG54843.1; GSPDB:GN00145; UWGP:206
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0639

Query Match 8.2%; Score 109.5; DB 2; Length 338;
Best Local Similarity 20.1%; Pred. No. 0.26;
Matches 56; Conservative 32; Mismatches 92; Indels 99; Gaps 9;

QY 73 TYLELQHYMPYTPVLVTYAPGVSPSP-----ILLYPMSDDP 108
DB 65 TYLTEHHYIPSETPTTQTPTVQTPDAGQKTAATGDVQTTARYQSMINAROSAVTDAQ 124

QY 109 QLGINRQQLK-----NLYSY-----FNDLR 129
DB 125 OTQITTEQQAIVATQKTLAATGDTQNTAHYQEMINARLAQNEANQRTATEQGGKNALT 184

QY 130 HDFKLKVLDAISKNKQ-----NIDTISKYLLLELGYLDDSYRMMEQNTHNI 176
DB 185 TDVAVQQONERTQYDKMQSLAQESAQAHEQIDLSQDVQTTQHQLTNTQKRVADNSQOI 244

QY 177 NKLS-----KELOTGLANQSALSMVQPNVGVKTS-----VSAAGVGRDKT 218
DB 245 NTLNNHFFSLKNEVDNRRKEANAGTA--SAIAIASQPQ--VKTGDMVMVMSAGAGTFNGES 300

QY 219 ALAIGVGSRTITDFTAKAGVAFNTYNGMSYGASVGYEF 257

Db 301 AVSGTGFNFAGTHTVLKAGISADT-QSDFGAGVGVGSF 338

RESULT 10
AF0169
probable exported protein YP01388 [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AF0169
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90217.1; PID:gl5979437; GSPDB:GNO0175
C:GeneID:8
A:Gene: YP01388

Query Match 7.6%; Score 102.5; DB 2; Length 622;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 37; Conservative 31; Mismatches 63; Indels 17; Gaps 6;

Qy 113 NRQQLKLNLYSFNDLRHDFKLVLDARISKKNQIDT-ISKYLL-----ELGTLYLDDSYR 167
Db 485 NLIDVTLLNANNYTDLS-----VNTIITYTGKQYTDSDRINEYQRTFKNEFLTYSNGKRF 537
Qy 168 MMEQNTNINKLSKELDTGLANQSALSMVLQPNVGKTSVAAGVGYRDKTALAIAGVGR 227
Db 538 GFDRK---DINQKQQLNAGIAATWAAVTPQKSG-SKVSIGVGLAGYSQDAGSGVGAIW 593
Qy 228 ITRFTAKAGVAFNTYNG-QMSYGASVG 254
Db 594 VNQRITWNTTVDYTGQVSLTGLSIG 621

RESULT 11
S76235
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76235
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-831 <KAN>
A:Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAA18494.1; PID:di01922
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Keywords: ATP; nucleotide binding; p-loop
F:317-324/Region: nucleotide-binding motif A (p-loop)
F:409-414/Region: nucleotide-binding motif B
F:413-416/Region: DEXH motif

Query Match 7.4%; Score 99.5; DB 2; Length 831;
Best Local Similarity 20.1%; Pred. No. 5.8;
Matches 52; Conservative 32; Mismatches 70; Indels 105; Gaps 9;

Qy 48 SNEGGFDI-----KVPGIKMKPREWISKQ-ATYLE-----LQHYMPYTPVLVTVYAGVSP 96
Db 167 SKEGDYSYRLQRKQVGLTKALRDLLKRWVESYLEPPKAVQPOLPKQWTLNFDP---- 222
Qy 97 SPILLYPWSPDQGLGINROQLKLNLYSFNDLRHDFKLVLDAR---ISKKNQIDTISK 153

Db 223 -----SFKLTQLQARTLIVLKNQGG----- 242

Qy 154 YLLELGTLYLDDSYRMEQNTNINKLSKELDTGLANQSALSMVLQPN----- 200

Db 243 -----EMWLADLLKAVPCSASTIOSLAKKGLVLAIREKRLRFOQPSINASOAPELTPAQ 297

Qy 201 -----GVGKTSVSAAGGYR---DKTALA-----ICVGSRIITDR 231

Db 298 KQACQTVLPLOGYHQVLLHGVYGGSKTEVYLIQICDRLRGKQSVLVLVPELGLTPQLTDR 357

Qy 232 FTAKAGVAFNTYNGMSY 250

Db 358 FRARFGNKVAVVYTHGLSSG 376

RESULT 12
A56921
kinesin family protein KIF1a - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C:Accession: A56921
R:Okada, Y.; Yamazaki, H.; Sekine-Aizawa, Y.; Hirokawa, N. Cell 81, 769-780, 1995
A:Title: The neuron-specific kinesin superfamily protein KIF1a is a unique monomeric motor
A:Reference number: A56921; MUID:95292344; PMID:7539720
A:Accession: A56921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1695 <RES>
A:Cross-references: GB:D29951; NID:g976234; PIDN:BAA06221.1; PID:g976235
C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; pleckstrin
C:Keywords: nucleotide binding; p-loop
F:6-360/Domain: kinesin motor domain homology <XMOT>
F:97-104/Region: nucleotide-binding motif A (p-loop)

Query Match 7.3%; Score 97.5; DB 2; Length 1695;
Best Local Similarity 24.1%; Pred. No. 23;
Matches 47; Conservative 34; Mismatches 71; Indels 43; Gaps 10;

Qy 4 KCLVAVVGLACSTTTTMAQQPKFAGVSSLYSYEYDYGKRWTSNEGDFDKVPGIKMK 63
Db 25 KCIIQMSG---STTIVNPKQPKETPKS--FSFDYSY-----MSHTSPEDI----- 65

Qy 64 PKEMLSKQATYLE-----LQH-YMPYTPVLVTVYAGVSPSPILLYPMSDDPOLGINROQL 117

Db 66 --NYASQKVYRDIGEEMLQHAFGYGVNVCIFAYGTGAGKSYTMGMKQKQGGI-IFQL 122

Qy 118 KLNLYSYFNDLRHD-----FKLVLDARISKKNQIDTISKYLLLELGTLYLDD 164

Db 123 CEDLESRINDTNDNMSYSVEVSMEIYCERVDLLNPKNGNL-RVREHPL-LGPIYVED 180

Qy 165 SYRMEQNTNINKL 179

Db 181 LSKLAVTSYNDIQDL 195

RESULT 13
A95179
hypothetical protein SPI538 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: A95179
R:Petzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidon, J.D.; Unayan, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, nson, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95179
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-466 <KUR>
A;Cross-references: GB:AF005672; PIDN:AAK75626.1; PID:gl4973027; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1538

Query Match 7.2%; Score 97; DB 2; Length 466;
Best Local Similarity 24.7%; Pred. No. 4.2;
Matches 59; Conservative 30; Mismatches 88; Indels 62; Gaps 14;
Qy 45 WTSNEGFDIKVP---GIKMKPEKWKISKOATYLELQHYMPYTPVLVYAPGV----- 94
Db 156 WTFEDKGD-DLHLPDSLSDKLRVW---HQHSSDIVPISGSKATGVEKVVVHL 205
Qy 95 --SPSPILLYPMSPDPQLGINROQLKLNLYSYFN-----DLRHDFKLVLDARISKKNQN 147
Db 206 GLKPEKVMVF--GD-----GLN-----DLELFDYAGISVAMGISHD-----KIKEK 244
Qy 148 IDTISKVLELGTLYD-DSYRMEQNTN-----NINKLSKELQTLGLANQSALSMVLQVNGV 202
Db 245 ADYITKLEEDGIFPAALFVGMVEKELHFPQVDIETVEGPLATIKTNHGDLRIKLFPEHA 304
Qy 203 GKT---SVSAAVGGRDKTALAIGVSRITDRFTAKAGVAFNTYNGGMS-YGASVGVYEF 257
Db 305 PKTVANFVLSKDGYYDGV-----IFHRIIKDFMIQGGDPTGTGMGGESYIGESFEDEF 358

RESULT 14
D64226
N;Endopeptidase La (EC 3.4.21.53) - Mycoplasma genitalium
N;Alternate names: ATP-dependent proteinase lon; ATP-dependent serine proteinase La
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
C;Species: Mycoplasma genitalium
C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Jun-2003
C;Accession: D64226
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346; PMID:7569993
A;Accession: D64226
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-795 <TIGR>
A;Cross-references: GB:U39702; GB:I43967; NID:g1045928; PID:g1045929; TIGR:MG239
A;Experimental source: strain G-37
C;Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP
C;Genetics:
A;Gene: lon
A;Genetic code: SGC3
C;Superfamily: ATP-dependent Lon protease
C;Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine
F;379-386/Region: nucleotide-binding motif A (P-loop)
F;442-447/Region: nucleotide-binding motif B
F;702/Active site: Ser #status predicted

Query Match 7.2%; Score 96.5; DB 1; Length 795;
Best Local Similarity 22.4%; Pred. No. 9.6;
Matches 49; Conservative 39; Mismatches 90; Indels 41; Gaps 8;
Qy 28 AGVSSLYSEYDYKGGKWTWSNEGFDIKVPGIGKMKPEKWKISKOATYLELQHYMPYTPVL 87
Db 208 AGIEDLQIQ-DYGRSK---NKETFDKLSKTRKINEQLSRQ---QRDFVLRKLR 259
Qy 88 VTYAPGVSPSPILLYPMSPDPQLGINROQLKLNLY-----SYFNDLRHDFKLVLDARI 141
Db 260 IREEIGIS-----SKKEDVASIRKKLDENPYPAIKKRILSELEHYENSSSSQES 311
Qy 142 SKKNQIDTISKVLEL-----GTYLDDSYRMEQNTNINKLSKELQTLGLA----- 188
Db 312 TLTKTYIDT-----LLNLPMWQSKSDNSDVKNLIKTLDKNHTGLDKVKERIVEILAVQLRT 367

Qy 189 --NQSALSMVLQVNGVKTSSVSAAVGGYRDKTALAIGVG 225
Db 368 QKNKGPIINCLVGPVGVKSLAKSIABALDKKFKVKSIG 406

RESULT 15

H98045
hypothetical protein sprl393 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: H98045
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: H98045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-466 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAI00197.1; PID:g15459043; GSPDB:GN00174
C;Genetics:
A;Gene: sprl393

Query Match 7.1%; Score 95; DB 2; Length 466;
Best Local Similarity 24.7%; Pred. No. 6.1;
Matches 59; Conservative 30; Mismatches 88; Indels 62; Gaps 14;
Qy 45 WTSNEGFDIKVP---GIKMKPEKWKISKOATYLELQHYMPYTPVLVYAPGV----- 94
Db 156 WTFEDKGD-DLHLPDSLSDKLRVW---HQHSSDIVPISGSKATGVEKVVVHL 205
Qy 95 --SPSPILLYPMSPDPQLGINROQLKLNLYSYFN-----DLRHDFKLVLDARISKKNQN 147
Db 206 GLKPEKVMVF--GD-----GLN-----DLELFDYAGISVAMGISHD-----KIKEK 244
Qy 148 IDTISKVLELGTYY-LDDSYRMEQNTN-----NINKLSKELQTLGLANQSALSMVLQVNGV 202
Db 245 ADYITKLEEDGIFDALEFVGMVEKELHFPQVDIETVEGPLATIKTNHGDLRIKLFPEHA 304
Qy 203 GKT---SVSAAVGGRDKTALAIGVSRITDRFTAKAGVAFNTYNGGMS-YGASVGVYEF 257
Db 305 PKTVANFVLSKDGYYDGV-----IFHRIIKDFMIQGGDPTGTGMGGESYIGESFEDEF 358

Search completed: May 13, 2004, 08:16:00
Job time : 22 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 17 Seconds
(without alignments)
787.178 Million cell updates/sec

Title: US-10-030-529a-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFNTYNGMSYGASGVGEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	186.5	13.9	455	1	YADA YEREN
2	182	13.6	434	1	YADA YEREN
3	99.5	7.4	831	1	PRIA_SYNY3
4	99	7.4	1690	1	KPIA_HUMAN
5	97.5	7.3	1695	1	KPIA_MOUSE
6	96.5	7.2	1795	1	LON_MYCGE
7	92.5	6.9	588	1	CMC2_CABEL
8	92.5	6.9	727	1	MPPI_ARATH
9	90.5	6.7	819	1	LON_CHLPN
10	90.5	6.7	2716	1	OSA_DROME
11	90	6.7	1328	1	POLX_TOBAC
12	89.5	6.7	956	1	UVRA_LISIN
13	88.5	6.6	498	1	FLID_BACSU
14	88	6.6	4543	1	LRPI_CHICK
15	87	6.5	806	1	LON_FORBU
16	86.5	6.4	524	1	CLPP_CHURE
17	86.5	6.4	1226	1	DSRA_HUMAN
18	86	6.4	473	1	Y085_UREPA
19	86	6.4	766	1	TKL1_MOUSE
20	86	6.4	920	1	YML1_YEAST
21	85.5	6.4	218	1	GTH5_ARATH
22	85.5	6.4	522	1	LEU1_SHEON
23	85	6.3	677	1	YD64_MYCPN
24	85	6.3	810	1	HPUB_NEIMC
25	84	6.3	491	1	GEPD_BUCAI
26	84	6.3	662	1	YME1_SCHMA
27	84	6.3	1011	1	PERQ_MOUSE
28	83.5	6.2	763	1	PEPX_STRMC
29	83.5	6.2	817	1	PERQ_HUMAN
30	83.5	6.2	2214	1	POLG_CXA24
31	83	6.2	290	1	CYOA_BUCAP
32	83	6.2	518	1	GSHI_BUCAP
33	82.5	6.1	436	1	TOLB_RHINE

34	82.5	6.1	438	1	MURD_YERPE	Q8zif1 versinia pe
35	82.5	6.1	824	1	HELI_HSV6U	P52356 human herpe
36	82.5	6.1	824	1	HELI_HSV6Z	P52450 human herpe
37	82.5	6.1	956	1	UVRA_LISMO	Q8y4f6 listeria mo
38	82.5	6.1	1748	1	YNR2_YEAST	P53886 saccharomyc
39	82.5	6.1	1876	1	GLS1_YEAST	P38631 saccharomyc
40	81.5	6.1	313	1	ENG3_RHOBA	Q7uu26 rhodospirell
41	81.5	6.1	359	1	OMPA_SERMA	P04845 serratia ma
42	81.5	6.1	384	1	YZ43_METJA	Q60298 methanococc
43	81.5	6.1	826	1	CRAA_BACUH	Q96597 bacillus th
44	81	6.0	276	1	OCCT_AGRTU	P35121 agrobacteri
45	81	6.0	366	1	APNI_YEAST	P22936 saccharomyc

ALIGNMENTS

RESULT 1					
YADA YEREN					
ID	YADA_YEREN	STANDARD;	PRT;	455	AA.
AC	P31489;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Invasin precursor (Outer membrane adhesin).				
GN	YADA OR YOPA OR INVA OR YOP1.				
OS	Yersinia enterocolitica.				
OG	Plasmid pIV.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Yersinia.				
OX	NCBI_TaxID=630;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=6471/76 / Serotype O:3;				
RX	MEDLINE=35020586; PubMed=7934875;				
RA	Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,				
RA	Skurnik M.;				
RT	"Hydrophobic domains affect the collagen-binding specificity and				
RT	surface polymerization as well as the virulence potential of the Yada				
RT	protein of Yersinia enterocolitica.";				
RL	Mol. Microbiol. 10:995-1011(1993).				
CC	-1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO				
CC	PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE				
CC	CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS				
CC	PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL				
CC	SURFACE.				
CC	-1- SUBCELLULAR LOCATION: Outer membrane.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X13882; CAA32086.1; -.				
DR	PIR; S04912; S04912.				
DR	InterPro; IPR008126; Adhesion.				
DR	InterPro; IPR008640; Hep_Hag.				
DR	InterPro; IPR008635; HIM.				
DR	InterPro; IPR005594; Yada.				
DR	Pfam; PF05658; Hep_Hag; 4.				
DR	Pfam; PF05662; HIM; 1.				
DR	Pfam; PF03895; Yada; 1.				
DR	PRINTS; PR01756; OMADHESIN.				
DR	Plasmid; Virulence; Signal; Outer membrane.				
KW	Plasmid; Virulence; Signal; Outer membrane.				
FT	SIGNAL	1	25		
FT	CHAIN	26	455		
FT	INVASIN.				
SQ	SEQUENCE	455	AA;	47136	MW; AC12EF68C657DAC0 CRC64;

Query Match 13.9%; Score 186.5; DB 1; Length 455;
Best Local Similarity 34.1%; Pred. No. 6.3e-08;


```
Best Local Similarity 20.1%; Pred. No. 2;
Matches 52; Conservative 32; Mismatches 70; Indels 105; Gaps 9;

Qy 48 SNEGGFDI-----KVPGLKMPKWKISKQ--ATYLE-----LQHMYPYTPVLVTVAPGVSP 96
Db 167 SKEDGYSYRLQKVPGLTKALRDLKQWVSYLEPPKAVQPOLPKVMTLLNDFP-----222
Qy 97 SPILLYPMSPDQLGINRQOLKLNLYSYFNDLRHDFKLKVLDAK----ISKKNQONIDTISK 153
Db 223 -----SFKLTELOARTLVLKQGG-----242
Qy 154 YLLELGYLDSYRMMEQNTNINKSKELQTLGLANQSALSMVQPN-----200
Db 243 -----EWLADLLKAVPCSASTIQSLACKGLVAIAEREKRLFFQQPSINASQAPELTPAQ 297
Qy 201 -----GVGKTSVSAVGGYR---DKTALA-----IGVGSRTYDR 231
Db 298 KOACQTVLPQGYHQVLLHGVTSQKTEVLIQICDRLGKGQSVLVLVPEIGLTPQLTDR 357
Qy 232 FTAKAGVAFNTYNGMSYG 250
Db 358 FRFRGNKVAVTHSLSG 376

RESULT 4
KF1A HUMAN
ID_KF1A_HUMAN STANDARD; PRT; 1690 AA.
AC Q12756;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).
GN ATSV OR KIF1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96299637; PubMed=8661001;
RA Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;
RT "Characterization of a kinesin-related gene ATSV, within the tuberosus
RT sclerosis locus (TSC1) candidate region on chromosome 9q34.";
RL Genomics 33:421-429(1996).
CC -!- FUNCTION: Motor for anterograde axonal transport of synaptic
CC vesicle precursors (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104
CC subfamily.
CC -!- SIMILARITY: Contains 1 FHA domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90840; CAA62346.1; +
CC HSSP; P17119; 3KAR.
CC Genew; HGNC:888; ATSV.
CC MIW; 601255; -.
CC GO; GO:0003774; F:motor activity; TAS.
CC GO; GO:0008089; P:anterograde axon cargo transport; TAS.
CC InterPro; IPR000253; FHA.
CC InterPro; IPR001752; kinesin_motor.
CC InterPro; IPR001849; PH.
CC InterPro; IPR008984; SMAD_FHA.
CC Pfam; PF00498; FHA; 1.
CC Pfam; PF00225; Kinesin; 1.
```

```
Pfam; PF00169; PH; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00129; Kisc; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR PROSITE; PS0006; FHA DOMAIN; 1.
DR PROSITE; PS0003; PH DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 361 KINESIN-MOTOR.
FT DOMAIN 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).
FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1575 1673 PH.
FT NP BIND 97 104 ATP (POTENTIAL).
SQ SEQUENCE 1690 AA; 191083 MW; D8DDEC784624FB4D CRC64;

Query Match 7.4%; Score 99; DB 1; Length 1690;
Best Local Similarity 24.1%; Pred. No. 5.5;
Matches 53; Conservative 34; Mismatches 83; Indels 50; Gaps 11;

Qy 4 KCLVAVVGLACSTTTTMAQQPPKPAQVSSLYSYEYDYGKWKWTWSNEGGFDIKVPGIKMK 63
Db 25 KCIQMSG---STTIYNPKQKPTPKS--FSPDY- ----WSHTSPEDI-----65
Qy 64 PKWISKOATYLE-----LQH-YMPYTPVLVTVAPGVSPSPILLYPMSPDQLGINRQOL 117
Db 66 --NYASQKVYRDIGEMQLHAFEGYVNCIFAYGTGAGKSYTMMGKQKQOGI-IPQL 122
Qy 118 KLNLYSFPNLRHD-----FRLKVLDAISKKNQONIDTISKYLELLEGTLYDD 164
Db 123 CEDLFSINDTNDNMSYSEVSYMEYICERVRDLNPKNGNL-RVREHPL-LGPYVED 180
Qy 165 -----SYRMQEQTNTINKSKELQTLGLANQSALSMV 197
Db 181 LSKLAVTSYNDIQDLMDSGNKARTVAATNNWETSRSRAV 220

RESULT 5
KF1A MOUSE
ID_KF1A_MOUSE STANDARD; PRT; 1695 AA.
AC P31773; Q61770;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).
GN ATSV OR KIF1A OR KIF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95292344; PubMed=7539720;
RA Okada Y., Yamazaki H., Sekine-Aizawa Y., Hirokawa N.;
RT "The neuron-specific kinesin superfamily protein KIF1A is a unique
RT monomeric motor for anterograde axonal transport of synaptic vesicle
RT precursors.";
RL Cell 81:769-780(1995).
RN [2]
RP PRELIMINARY SEQUENCE OF 100-247 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RL J. Cell Biol. 119:1287-1296(1992).
CC -!- FUNCTION: Motor for anterograde axonal transport of synaptic
CC vesicle precursors.
```

CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
CC TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE
CC TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE
CC CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE
CC PERINUCLEAR AND SYNAPTIC REGIONS.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104
CC subfamily.
CC -1- SIMILARITY: Contains 1 FHA domain.
CC -1- SIMILARITY: Contains 1 PH domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D29951; BAA06221.1; --
CC PIR; A56921; A56921.
CC DR PDB; 1I58; 30-MAY-01.
CC DR PDB; 1I61; 30-MAY-01.
CC DR MGD; MGI:108391; Kif1a.
CC DR InterPro; IPR000253; FHA.
CC DR InterPro; IPR001752; kinesin_motor.
CC DR InterPro; IPR001849; PH.
CC DR InterPro; IPR008984; SNAD_FHA.
CC DR Pfam; PF00498; FHA; 1.
CC DR Pfam; PF00225; kinesin; 1.
CC DR Pfam; PF00169; PH; 1.
CC DR PRINTS; PR00380; KINESINHEAVY.
CC DR SMART; SM00240; FHA; 1.
CC DR SMART; SM00129; KISC; 1.
CC DR SMART; SM00233; PH; 1.
CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
CC DR PROSITE; PS50067; FHA DOMAIN; 1.
CC DR PROSITE; PS50003; PH DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; 3D-structure.
FT DOMAIN 1 361 KINESIN-MOTOR.
FT DOMAIN 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).
FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1580 1678 PH.
FT NP BIND 97 104 ATP (BY SIMILARITY).
SQ SEQUENCE 1695 AA; 191724 MW; D6EC3B88CBC9CC6 CRC64;

Query Match 7.3%; Score 97.5; DB 1; Length 1695;
Best Local Similarity 24.1%; Pred. No. 7.3;
Matches 47; Conservative 34; Mismatches 71; Indels 43; Gaps 10;

Qy 4 KCLVAVVGLACSTTTMAQQPKFAGVSSLYSEYDYGKGKWTWSNEGDFDIKVPGRMK 63
Db 25 KCIQMSG---STTTINPRQPKETPKS--FSPDYSY-----WSHSPEDI----- 65

Qy 64 KPEWISQKATYLE-----LQH-YMPYTPVLVYAPGVSPSILLYPMSDPDQLGINRQQL 117
Db 66 --NVASQKVYRDIGEEMLQHAPEGVNVCIFAYQGTGAGKSYTMGMGKQKDDQGI-IPQL 122

Qy 118 KNLKLYSFNDLRHP-----FKLKVLDAISKKNQKIDTTISKVLLLEGTGLDD 164
Db 123 CEDLFSRINDTNDNMSYSEVSYNEICERVRLNLPKNGNL-RVREHPL-LGPYVD 180

Qy 165 SYRMMEQNTHNINKL 179
Db 181 LSKLAVTSYNDIQDL 195

RESULT 6
LON_MYCGE

ID LON_MYCGE STANDARD; PRT; 795 AA.
AC P47481; Q49276;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR MG219.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Frichman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL Science 270:397-403 (1995).
RN [2]
RP SEQUENCE OF 484-606 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing";
RL J. Bacteriol. 175:7918-7930 (1993).
CC -1- FUNCTION: DEGRADATES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADATES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39703; AAC71460.1; --
CC EMBL; U02148; AAD12428.1; --
CC PIR; D64226; D64226.
CC MEROPS; S16.004; --
CC TIGR; MG239; --
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003959; AAA ATPase centr.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; Peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMs; TIGR00763; Lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT DOMAIN 304 308 POLY-SER.
FT NP BIND 379 386 ATP (POTENTIAL).
FT ACT_SITE 702 702 BY SIMILARITY.
FT CONFLICT 494 494 T -> S (IN REF. 2).

FT	TRANSMEM	307	326	2 (POTENTIAL).
FT	TRANSMEM	352	365	3 (POTENTIAL).
FT	TRANSMEM	403	422	4 (POTENTIAL).
FT	TRANSMEM	446	463	5 (POTENTIAL).
FT	TRANSMEM	504	523	6 (POTENTIAL).
FT	CA_BIND	52	63	EF-HAND 1.
FT	CA_BIND	86	97	EF-HAND 2.
FT	DOMAIN	123	133	ANCESTRAL CALCIUM SITE 3.
FT	CA_BIND	153	164	EF-HAND 4.
FT	REPEAT	246	332	SOLCAR 1.
FT	REPEAT	342	428	SOLCAR 2.
FT	REPEAT	440	529	SOLCAR 3.
SEQ	SEQUENCE	588 AA;	66336 MW;	68DDF60923D8697D CRC64;

Query Match 6.9%; Score 92.5; DB 1; Length 588;
Best Local Similarity 22.4%; Pred. No. 5;
Matches 48; Conservative 41; Mismatches 72; Indels 53; Gaps 12

QY	49	NEGGFDIKVGIKMKPKEWISKQATYLELQHYMPYTPVLVYTAGVSPSPILLVPMSPDP	108
Db <td>89<td>NDGTIDIR-----DLTLAKHETHIP-----ANLAPVIMSKVS-PD</td><td>124</td></td>	89 <td>NDGTIDIR-----DLTLAKHETHIP-----ANLAPVIMSKVS-PD</td> <td>124</td>	NDGTIDIR-----DLTLAKHETHIP-----ANLAPVIMSKVS-PD	124
QY <td>109<td>QLGINRQQLKLNLYSYFN-DLRHDFKLKVLDRISKKNQN-IDTI--SKYLLELGTYYLDD</td><td>164</td></td>	109 <td>QLGINRQQLKLNLYSYFN-DLRHDFKLKVLDRISKKNQN-IDTI--SKYLLELGTYYLDD</td> <td>164</td>	QLGINRQQLKLNLYSYFN-DLRHDFKLKVLDRISKKNQN-IDTI--SKYLLELGTYYLDD	164
Db <td>125<td>DEG-----RVDYFSSSYYLENEQKLAEMFADMRNHDGLVDVVMKMYCKDIGVPLDD</td><td>178</td></td>	125 <td>DEG-----RVDYFSSSYYLENEQKLAEMFADMRNHDGLVDVVMKMYCKDIGVPLDD</td> <td>178</td>	DEG-----RVDYFSSSYYLENEQKLAEMFADMRNHDGLVDVVMKMYCKDIGVPLDD	178
QY <td>165<td>SYRMEQNTNINKLSKELQTLGANSALS-----MLVQPNGVGKTSVSAAGVGYRDKTAL</td><td>220</td></td>	165 <td>SYRMEQNTNINKLSKELQTLGANSALS-----MLVQPNGVGKTSVSAAGVGYRDKTAL</td> <td>220</td>	SYRMEQNTNINKLSKELQTLGANSALS-----MLVQPNGVGKTSVSAAGVGYRDKTAL	220
Db <td>179<td>-----HKAQHVNMND---QTGSASVDLKEFQEFMMLYPS-----SDLKDIYDFVFRHNLI</td><td>226</td></td>	179 <td>-----HKAQHVNMND---QTGSASVDLKEFQEFMMLYPS-----SDLKDIYDFVFRHNLI</td> <td>226</td>	-----HKAQHVNMND---QTGSASVDLKEFQEFMMLYPS-----SDLKDIYDFVFRHNLI	226
QY <td>221<td>AIGVGRITDRFT---AKAGVAFTNYNGMSYGA</td><td>251</td></td>	221 <td>AIGVGRITDRFT---AKAGVAFTNYNGMSYGA</td> <td>251</td>	AIGVGRITDRFT---AKAGVAFTNYNGMSYGA	251
Db <td>227<td>DIGEDSQIPEDFSQEQMOEGIMWRHLVAGGAAGA</td><td>260</td></td>	227 <td>DIGEDSQIPEDFSQEQMOEGIMWRHLVAGGAAGA</td> <td>260</td>	DIGEDSQIPEDFSQEQMOEGIMWRHLVAGGAAGA	260

RESULT 8

MFPI_ARATH	STANDARD;	PRT;	727 AA.
AC	Q9LW85;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	MAR binding filament-like protein 1.		
GN	MFPI OR AT3G16000 OR MSL1.4.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid 1; Brassicales; Brassicaceae; Arabidopsids.		
NCBI_TaxID=3702;			
[1]	SEQUENCE FROM N.A.		
RP	STRAIN=cv. Columbia;		
RC	MEDLINE=20277480; PubMed=10819329;		
RA	Sato S., Nakamura Y., Kaneo T., Katoh T., Asamizu E., Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence		
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC		
RT	clones.";		
RT	DNA Res. 7:131-135(2000).		
RL	-!- FUNCTION: Binds DNA. Interacts with chromatin via matrix		
CC	attachment regions (MARS). Likely to participate in nuclear		
CC	architecture by connecting chromatin with the nuclear matrix and		
CC	potentially with the nuclear envelope (By similarity).		
CC	-!- SUBCELLULAR LOCATION: Nuclear matrix (By similarity).		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the-European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; AB012247; BAB02666.1; -.		

RESULT 10
 ID_ DROME STANDARD; PRT; 2716 AA.
 AC OSIN94; O61603; Q9VEG7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Trithorax group protein OSA (Eyelid protein).
 GN OSA OR ELD OR CG7467.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
 RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=97415319; PubMed=9271118;
 RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
 RT "eyelid antagonizes wingless signaling during Drosophila development
 and has homology to the Bright family of DNA-binding proteins.";
 RL Genes Dev. 11:1949-1962(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Berkley.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.E., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.K., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.M., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.D., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D.A., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP FUNCTION, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=99112962; PubMed=9895321;
 RA Vazquez M., Moore L., Kennison J.A.;
 RT "The trithorax group gene osa encodes an ARID-domain protein that
 genetically interacts with the brahma chromatin-remodeling factor to
 regulate transcription";
 RL Development 126:733-742(1999).

[4]
 RNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
 RX MEDLINE=20069333; PubMed=10601025;
 RA Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
 RT "Osa associates with the Brahma chromatin remodeling complex and
 promotes the activation of some target genes";
 RL EMBO J. 18:7029-7040(1999).
 RN [5]
 RP FUNCTION
 RX MEDLINE=99403006; PubMed=10471712;
 RA Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
 RT "A genetic screen for modifiers of E2F in Drosophila melanogaster";
 RL Genetics 153:275-287(1999).
 RN [6]
 RP IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
 RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
 RX MEDLINE=20270023; PubMed=10809665;
 RA Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
 RT "The Drosophila brahma complex is an essential coactivator for the
 trithorax group protein zeste";
 RL Genes Dev. 14:1058-1071(2000).
 RN [7]
 RP FUNCTION AS A COREPRESSOR.
 RX MEDLINE=20573925; PubMed=11124806;
 RA Collins R.T., Treisman J.E.;
 RT "Osa-containing Brahma chromatin remodeling complexes are required
 for the repression of wingless target genes";
 RL Genes Dev. 14:3140-3152(2000).
 RN [8]
 RP FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
 RX MEDLINE=22515897; PubMed=12629041;
 RA Heitzler P., Vanollet L., Biryukova I., Romain P.;
 RT "Enhancer-promoter communication mediated by Chip during
 Pannier-driven proneural patterning is regulated by Osa";
 RL Genes Dev. 17:591-596(2003).
 CC -!- FUNCTION: Trithorax group (trxG) protein required for embryonic
 segmentation, development of the notum and wing margin, and
 photoreceptor differentiation. Required for the activation of
 genes such as Antp, Ubx and Eve. Binds to DNA without specific
 affinity, suggesting that it is recruited to promoters by
 promoter-specific proteins. Essential component of the Brahma
 complex, a multiprotein complex which is the equivalent of the
 yeast SWI/SNF complex and acts by remodelling the chromatin by
 catalyzing an ATP-dependent alteration in the structure of
 nucleosomal DNA. This complex can both serve as a transcriptional
 coactivator or corepressor, depending on the context. Acts as an
 essential coactivator for Zeste, which recruits the whole complex
 to specific genes. In contrast, it acts as a corepressor for Wg
 target genes, possibly via an interaction with Pan and Gro. It
 also acts as a negative regulator for proneural achaete-scute,
 when it is directly recruited by Pan and Chi. Also represses E2f
 activation.
 CC -!- SUBUNIT: Component of the Brahma complex, which is composed of
 Brm, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
 CC Interacts with Pnr and Chi via its EHD domain.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
 third instar larvae, it is ubiquitously expressed in wing and eye-
 antenna imaginal disks, with a stronger expression in a band just
 anterior to the morphogenetic furrow.
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -!- DOMAIN: The ARID domains mediates the binding to DNA.
 CC -!- SIMILARITY: Contains 1 ARID domain.
 CC -!- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
 CC -!- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
 erroneous gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```
or send an email to license@iab-sib.ch).
-----
CC EMBL; AF053091; AAC06254.1; -.
DR EMBL; AE003718; AAF55457.1; ALT_SEQ.
DR EMBL; AE003718; AAN13750.1; -.
DR PIR; T13049; T13049.
DR FlyBase; FBgn0030313; osa.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
DR GO; GO:0045449; P:regulation of transcription; IDA.
DR GO; GO:0007379; P:segment specification; IMP.
DR GO; GO:0008587; P:wing margin morphogenesis; IMP.
DR GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.
DR InterPro; IPR001506; ARID.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR006031; XYPPX.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02162; XYPPX; 8.
DR SMART; SM00501; BRIGHT; 1.
KW Transcription regulation; DNA-binding; Activator; Repressor;
KW Chromatin regulator; Nuclear protein; Developmental protein.
FT DOMAIN 997 1111
FT DOMAIN 1769 2517
FT DOMAIN 19 1763 PRO-RICH.
FT DOMAIN 174 380 GLN-RICH.
FT DOMAIN 619 873 GLY-RICH.
FT DOMAIN 1222 1453 GLY-RICH.
FT DOMAIN 1271 1751 GLN-RICH.
FT DOMAIN 1730 1745 HIS-RICH.
FT DOMAIN 2589 2624 SER-RICH.
FT DOMAIN 2625 2716 ALA-RICH.
FT CONFLICT 61 61 MISSING (IN REF. 1).
FT CONFLICT 1169 1169 V -> G (IN REF. 1).
FT CONFLICT 1795 1795 M -> T (IN REF. 1).
FT CONFLICT 2637 2637 G -> E (IN REF. 1).
SQ SEQUENCE 2716 AA; 284063 MW; BFAE76CB51C7C675 CRC64;

Query Match 6.7%; Score 90.5; DB 1; Length 2716;
Best Local Similarity 23.1%; Pred. No. 50;
Matches 46; Conservative 24; Mismatches 86; Indels 43; Gaps 8;

QY 67 WISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPQLGINRQQLNLYSYFN 126
Db 1006 WLDKLRAPWEER-----TP--ITACTISKQPLDIY-----RLIYVK 1042

QY 127 DLRHDFLKLVDARISKKNQIDITISKYLELGTLYDDSYRMEQNTNINKLSKELOTG 186
Db 1043 E-RGGF-----VEVTKSKTWKDIAG--LLGIGASSAAAYTLRKHYYTKNLLTFECHFDKG 1093

QY 187 LANQSALSMVQPNVGKTSVSAV-----GGYRDKTALAIGVGSRTIDRFTAKAGVAFNT 242
Db 1094 DIDPLFIQQVEAGSKKTKAKAASVPSPGGHLDAGTTNSTGSSNSQDSFPAPPGSAPNA 1153

QY 243 -----YNGGMSYGASVG 254
Db 1154 AIDGPGPGGSPYPVAGS 1172

RESULT 11
POLX TOBAC
ID POLX TOBAC STANDARD; PRT; 1328 AA.
AC P10978;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Retrovirus-related Pol polyprotein from transposon TNT 1-94 [Contains:
DE Protease [EC 3.4.23.-]; Reverse transcriptase [EC 2.7.7.49];
DE Endonuclease].
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiids; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxID=4097;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=89097311; PubMed=2536143;
RA Grandbastien M.-A., Spielmann A., Caboche M.;
RT "Tnt1, a mobile retroviral-like transposable element of tobacco
RL Nature 337:376-380(1989).
CC -! CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC -! (DNA) (N).
CC -! SIMILARITY: HIGH, WITH DROSOPHILA COPIA ELEMENT.
CC -! SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
CC -! SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@iab-sib.ch)
CC -----
DR EMBL; X13777; CAA32025.1; -.
DR PIR; S04273; S04273.
DR MEROPS; A11.002; -.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00158; ZF_CCHC; 1.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KW Endonuclease; Transferase; Polyprotein; Transposable element;
KW Zinc-finger.
FT ZN_FING 230 247 CCHC-TYPE.
FT ACT_SITE 297 297 PROTEASE (BY SIMILARITY).
SQ SEQUENCE 1328 AA; 151076 MW; F27E76C504B19B1B CRC64;

Query Match 6.7%; Score 90; DB 1; Length 1328;
Best Local Similarity 24.0%; Pred. No. 22;
Matches 47; Conservative 31; Mismatches 78; Indels 40; Gaps 9;

QY 59 GIKMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDPQLGINRQ-- 116
Db 962 GLKQAPRWYMKFDSFMKSTYLTYSDCPYFKRFSENNFIILLYYDDMLIVGDKGL 1021

QY 117 -LKLNLYSYFNDLRHDFLK-----VLDARISKKNQIDITISKYLELGTLYDDSYRM 168
Db 1022 IAKLK-----GDLSEKDFMDKDLGPAQLLGKVIKRIER-----TSRKLWLSQEKYIE---RV 1069

QY 169 MEQ-NTNINKLSKELOTGLANQSALSMVQPNVG-----GKTSVSAVAGGY----- 214
Db 1070 LERENMKN---AKPVSTPLAGHLKLSKMKCPTTVEERKGNMAKVPYSSAVGSLMYAVCT 1125

QY 215 RDKTALAIGVGSRTID 230
Db 1126 RPDIAHAGVVSVRFLE 1141

RESULT 12
UVRA LISIN
ID UVRA LISIN STANDARD; PRT; 956 AA.
AC Q928A5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UVABC system protein A (UvrA protein) (Excinuclease ABC subunit A).
GN Listeria OR Lin2631.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
```



```
FT DOMAIN 2472 2512 EGF-LIKE 10. 2512 2512 LDL-RECEPTOR CLASS A 11. 1116 1116 BY SIMILARITY.
FT DOMAIN 2516 2557 LDL-RECEPTOR CLASS A 12. 1129 1129 BY SIMILARITY.
FT DOMAIN 2558 2596 LDL-RECEPTOR CLASS A 13. 1138 1138 BY SIMILARITY.
FT DOMAIN 2597 2635 LDL-RECEPTOR CLASS A 14. 1157 1157 BY SIMILARITY.
FT DOMAIN 2636 2684 LDL-RECEPTOR CLASS A 15. 1170 1170 BY SIMILARITY.
FT DOMAIN 2688 2730 LDL-RECEPTOR CLASS A 16. 1180 1180 BY SIMILARITY.
FT DOMAIN 2730 2769 LDL-RECEPTOR CLASS A 17. 1194 1194 BY SIMILARITY.
FT DOMAIN 2770 2812 LDL-RECEPTOR CLASS A 18. 1204 1204 BY SIMILARITY.
FT DOMAIN 2814 2853 LDL-RECEPTOR CLASS A 19. 1219 1219 BY SIMILARITY.
FT DOMAIN 2854 2897 LDL-RECEPTOR CLASS A 20. 1235 1235 BY SIMILARITY.
FT DOMAIN 2900 2938 EGF-LIKE 11. 1244 1244 BY SIMILARITY.
FT DOMAIN 2939 2978 EGF-LIKE 12. 1259 1259 BY SIMILARITY.
FT DOMAIN 2979 3019 EGF-LIKE 13. 1551 1551 BY SIMILARITY.
FT DOMAIN 3020 3058 EGF-LIKE 14. 1561 1561 BY SIMILARITY.
FT DOMAIN 3059 3097 EGF-LIKE 15. 1576 1576 BY SIMILARITY.
FT DOMAIN 3098 3136 EGF-LIKE 16. 1857 1857 BY SIMILARITY.
FT DOMAIN 3137 3175 EGF-LIKE 17. 1867 1867 BY SIMILARITY.
FT DOMAIN 3176 3214 EGF-LIKE 18. 1882 1882 BY SIMILARITY.
FT DOMAIN 3215 3253 EGF-LIKE 19. 2166 2166 BY SIMILARITY.
FT DOMAIN 3254 3292 EGF-LIKE 20. 2176 2176 BY SIMILARITY.
FT DOMAIN 3293 3331 EGF-LIKE 21. 2190 2190 BY SIMILARITY.
FT DOMAIN 3332 3370 EGF-LIKE 22. 2487 2487 BY SIMILARITY.
FT DOMAIN 3371 3409 EGF-LIKE 23. 2497 2497 BY SIMILARITY.
FT DOMAIN 3410 3448 EGF-LIKE 24. 2511 2511 BY SIMILARITY.
FT DOMAIN 3449 3487 EGF-LIKE 25. 2531 2531 BY SIMILARITY.
FT DOMAIN 3488 3526 EGF-LIKE 26. 2544 2544 BY SIMILARITY.
FT DOMAIN 3527 3565 EGF-LIKE 27. 2555 2555 BY SIMILARITY.
FT DOMAIN 3566 3604 EGF-LIKE 28. 2572 2572 BY SIMILARITY.
FT DOMAIN 3605 3643 EGF-LIKE 29. 2585 2585 BY SIMILARITY.
FT DOMAIN 3644 3682 EGF-LIKE 30. 2597 2597 BY SIMILARITY.
FT DOMAIN 3683 3721 EGF-LIKE 31. 2609 2609 BY SIMILARITY.
FT DOMAIN 3722 3760 EGF-LIKE 32. 2621 2621 BY SIMILARITY.
FT DOMAIN 3761 3799 EGF-LIKE 33. 2633 2633 BY SIMILARITY.
FT DOMAIN 3800 3838 EGF-LIKE 34. 2645 2645 BY SIMILARITY.
FT DOMAIN 3839 3877 EGF-LIKE 35. 2657 2657 BY SIMILARITY.
FT DOMAIN 3878 3916 EGF-LIKE 36. 2669 2669 BY SIMILARITY.
FT DOMAIN 3917 3955 EGF-LIKE 37. 2681 2681 BY SIMILARITY.
FT DOMAIN 3956 3994 EGF-LIKE 38. 2693 2693 BY SIMILARITY.
FT DOMAIN 3995 4033 EGF-LIKE 39. 2705 2705 BY SIMILARITY.
FT DOMAIN 4034 4072 EGF-LIKE 40. 2717 2717 BY SIMILARITY.
FT DOMAIN 4073 4111 EGF-LIKE 41. 2729 2729 BY SIMILARITY.
FT DOMAIN 4112 4150 EGF-LIKE 42. 2741 2741 BY SIMILARITY.
FT DOMAIN 4151 4189 EGF-LIKE 43. 2753 2753 BY SIMILARITY.
FT DOMAIN 4190 4228 EGF-LIKE 44. 2765 2765 BY SIMILARITY.
FT DOMAIN 4229 4267 EGF-LIKE 45. 2777 2777 BY SIMILARITY.
FT DOMAIN 4268 4306 EGF-LIKE 46. 2789 2789 BY SIMILARITY.
FT DOMAIN 4307 4345 EGF-LIKE 47. 2801 2801 BY SIMILARITY.
FT DOMAIN 4346 4384 EGF-LIKE 48. 2813 2813 BY SIMILARITY.
FT DOMAIN 4385 4423 EGF-LIKE 49. 2825 2825 BY SIMILARITY.
FT DOMAIN 4424 4462 EGF-LIKE 50. 2837 2837 BY SIMILARITY.
FT DOMAIN 4463 4501 EGF-LIKE 51. 2849 2849 BY SIMILARITY.
FT DOMAIN 4502 4540 EGF-LIKE 52. 2861 2861 BY SIMILARITY.
FT DOMAIN 4541 4579 EGF-LIKE 53. 2873 2873 BY SIMILARITY.
FT DOMAIN 4580 4618 EGF-LIKE 54. 2885 2885 BY SIMILARITY.
FT DOMAIN 4619 4657 EGF-LIKE 55. 2897 2897 BY SIMILARITY.
FT DOMAIN 4658 4696 EGF-LIKE 56. 2909 2909 BY SIMILARITY.
FT DOMAIN 4697 4735 EGF-LIKE 57. 2921 2921 BY SIMILARITY.
FT DOMAIN 4736 4774 EGF-LIKE 58. 2933 2933 BY SIMILARITY.
FT DOMAIN 4775 4813 EGF-LIKE 59. 2945 2945 BY SIMILARITY.
FT DOMAIN 4814 4852 EGF-LIKE 60. 2957 2957 BY SIMILARITY.
FT DOMAIN 4853 4891 EGF-LIKE 61. 2969 2969 BY SIMILARITY.
FT DOMAIN 4892 4930 EGF-LIKE 62. 2981 2981 BY SIMILARITY.
FT DOMAIN 4931 4969 EGF-LIKE 63. 2993 2993 BY SIMILARITY.
FT DOMAIN 4970 5008 EGF-LIKE 64. 3005 3005 BY SIMILARITY.
FT DOMAIN 5009 5047 EGF-LIKE 65. 3017 3017 BY SIMILARITY.
FT DOMAIN 5048 5086 EGF-LIKE 66. 3029 3029 BY SIMILARITY.
FT DOMAIN 5087 5125 EGF-LIKE 67. 3041 3041 BY SIMILARITY.
FT DOMAIN 5126 5164 EGF-LIKE 68. 3053 3053 BY SIMILARITY.
FT DOMAIN 5165 5203 EGF-LIKE 69. 3065 3065 BY SIMILARITY.
FT DOMAIN 5204 5242 EGF-LIKE 70. 3077 3077 BY SIMILARITY.
FT DOMAIN 5243 5281 EGF-LIKE 71. 3089 3089 BY SIMILARITY.
FT DOMAIN 5282 5320 EGF-LIKE 72. 3101 3101 BY SIMILARITY.
FT DOMAIN 5321 5359 EGF-LIKE 73. 3113 3113 BY SIMILARITY.
FT DOMAIN 5360 5398 EGF-LIKE 74. 3125 3125 BY SIMILARITY.
FT DOMAIN 5399 5437 EGF-LIKE 75. 3137 3137 BY SIMILARITY.
FT DOMAIN 5438 5476 EGF-LIKE 76. 3149 3149 BY SIMILARITY.
FT DOMAIN 5477 5515 EGF-LIKE 77. 3161 3161 BY SIMILARITY.
FT DOMAIN 5516 5554 EGF-LIKE 78. 3173 3173 BY SIMILARITY.
FT DOMAIN 5555 5593 EGF-LIKE 79. 3185 3185 BY SIMILARITY.
FT DOMAIN 5594 5632 EGF-LIKE 80. 3197 3197 BY SIMILARITY.
FT DOMAIN 5633 5671 EGF-LIKE 81. 3209 3209 BY SIMILARITY.
FT DOMAIN 5672 5710 EGF-LIKE 82. 3221 3221 BY SIMILARITY.
FT DOMAIN 5711 5749 EGF-LIKE 83. 3233 3233 BY SIMILARITY.
FT DOMAIN 5750 5788 EGF-LIKE 84. 3245 3245 BY SIMILARITY.
FT DOMAIN 5789 5827 EGF-LIKE 85. 3257 3257 BY SIMILARITY.
FT DOMAIN 5828 5866 EGF-LIKE 86. 3269 3269 BY SIMILARITY.
FT DOMAIN 5867 5905 EGF-LIKE 87. 3281 3281 BY SIMILARITY.
FT DOMAIN 5906 5944 EGF-LIKE 88. 3293 3293 BY SIMILARITY.
FT DOMAIN 5945 5983 EGF-LIKE 89. 3305 3305 BY SIMILARITY.
FT DOMAIN 5984 6022 EGF-LIKE 90. 3317 3317 BY SIMILARITY.
FT DOMAIN 6023 6061 EGF-LIKE 91. 3329 3329 BY SIMILARITY.
FT DOMAIN 6062 6100 EGF-LIKE 92. 3341 3341 BY SIMILARITY.
FT DOMAIN 6101 6139 EGF-LIKE 93. 3353 3353 BY SIMILARITY.
FT DOMAIN 6140 6178 EGF-LIKE 94. 3365 3365 BY SIMILARITY.
FT DOMAIN 6179 6217 EGF-LIKE 95. 3377 3377 BY SIMILARITY.
FT DOMAIN 6218 6256 EGF-LIKE 96. 3389 3389 BY SIMILARITY.
FT DOMAIN 6257 6295 EGF-LIKE 97. 3401 3401 BY SIMILARITY.
FT DOMAIN 6296 6334 EGF-LIKE 98. 3413 3413 BY SIMILARITY.
FT DOMAIN 6335 6373 EGF-LIKE 99. 3425 3425 BY SIMILARITY.
FT DOMAIN 6374 6412 EGF-LIKE 100. 3437 3437 BY SIMILARITY.
FT DOMAIN 6413 6451 EGF-LIKE 101. 3449 3449 BY SIMILARITY.
FT DOMAIN 6452 6490 EGF-LIKE 102. 3461 3461 BY SIMILARITY.
FT DOMAIN 6491 6529 EGF-LIKE 103. 3473 3473 BY SIMILARITY.
FT DOMAIN 6530 6568 EGF-LIKE 104. 3485 3485 BY SIMILARITY.
FT DOMAIN 6569 6607 EGF-LIKE 105. 3497 3497 BY SIMILARITY.
FT DOMAIN 6608 6646 EGF-LIKE 106. 3509 3509 BY SIMILARITY.
FT DOMAIN 6647 6685 EGF-LIKE 107. 3521 3521 BY SIMILARITY.
FT DOMAIN 6686 6724 EGF-LIKE 108. 3533 3533 BY SIMILARITY.
FT DOMAIN 6725 6763 EGF-LIKE 109. 3545 3545 BY SIMILARITY.
FT DOMAIN 6764 6802 EGF-LIKE 110. 3557 3557 BY SIMILARITY.
FT DOMAIN 6803 6841 EGF-LIKE 111. 3569 3569 BY SIMILARITY.
FT DOMAIN 6842 6880 EGF-LIKE 112. 3581 3581 BY SIMILARITY.
FT DOMAIN 6881 6919 EGF-LIKE 113. 3593 3593 BY SIMILARITY.
FT DOMAIN 6920 6958 EGF-LIKE 114. 3605 3605 BY SIMILARITY.
FT DOMAIN 6959 6997 EGF-LIKE 115. 3617 3617 BY SIMILARITY.
FT DOMAIN 6998 7036 EGF-LIKE 116. 3629 3629 BY SIMILARITY.
FT DOMAIN 7037 7075 EGF-LIKE 117. 3641 3641 BY SIMILARITY.
FT DOMAIN 7076 7114 EGF-LIKE 118. 3653 3653 BY SIMILARITY.
FT DOMAIN 7115 7153 EGF-LIKE 119. 3665 3665 BY SIMILARITY.
FT DOMAIN 7154 7192 EGF-LIKE 120. 3677 3677 BY SIMILARITY.
FT DOMAIN 7193 7231 EGF-LIKE 121. 3689 3689 BY SIMILARITY.
FT DOMAIN 7232 7270 EGF-LIKE 122. 3701 3701 BY SIMILARITY.
FT DOMAIN 7271 7309 EGF-LIKE 123. 3713 3713 BY SIMILARITY.
FT DOMAIN 7310 7348 EGF-LIKE 124. 3725 3725 BY SIMILARITY.
FT DOMAIN 7349 7387 EGF-LIKE 125. 3737 3737 BY SIMILARITY.
FT DOMAIN 7388 7426 EGF-LIKE 126. 3749 3749 BY SIMILARITY.
FT DOMAIN 7427 7465 EGF-LIKE 127. 3761 3761 BY SIMILARITY.
FT DOMAIN 7466 7504 EGF-LIKE 128. 3773 3773 BY SIMILARITY.
FT DOMAIN 7505 7543 EGF-LIKE 129. 3785 3785 BY SIMILARITY.
FT DOMAIN 7544 7582 EGF-LIKE 130. 3797 3797 BY SIMILARITY.
FT DOMAIN 7583 7621 EGF-LIKE 131. 3809 3809 BY SIMILARITY.
FT DOMAIN 7622 7660 EGF-LIKE 132. 3821 3821 BY SIMILARITY.
FT DOMAIN 7661 7699 EGF-LIKE 133. 3833 3833 BY SIMILARITY.
FT DOMAIN 7700 7738 EGF-LIKE 134. 3845 3845 BY SIMILARITY.
FT DOMAIN 7739 7777 EGF-LIKE 135. 3857 3857 BY SIMILARITY.
FT DOMAIN 7778 7816 EGF-LIKE 136. 3869 3869 BY SIMILARITY.
FT DOMAIN 7817 7855 EGF-LIKE 137. 3881 3881 BY SIMILARITY.
FT DOMAIN 7856 7894 EGF-LIKE 138. 3893 3893 BY SIMILARITY.
FT DOMAIN 7895 7933 EGF-LIKE 139. 3905 3905 BY SIMILARITY.
FT DOMAIN 7934 7972 EGF-LIKE 140. 3917 3917 BY SIMILARITY.
FT DOMAIN 7973 8011 EGF-LIKE 141. 3929 3929 BY SIMILARITY.
FT DOMAIN 8012 8050 EGF-LIKE 142. 3941 3941 BY SIMILARITY.
FT DOMAIN 8051 8089 EGF-LIKE 143. 3953 3953 BY SIMILARITY.
FT DOMAIN 8090 8128 EGF-LIKE 144. 3965 3965 BY SIMILARITY.
FT DOMAIN 8129 8167 EGF-LIKE 145. 3977 3977 BY SIMILARITY.
FT DOMAIN 8168 8206 EGF-LIKE 146. 3989 3989 BY SIMILARITY.
FT DOMAIN 8207 8245 EGF-LIKE 147. 4001 4001 BY SIMILARITY.
FT DOMAIN 8246 8284 EGF-LIKE 148. 4013 4013 BY SIMILARITY.
FT DOMAIN 8285 8323 EGF-LIKE 149. 4025 4025 BY SIMILARITY.
FT DOMAIN 8324 8362 EGF-LIKE 150. 4037 4037 BY SIMILARITY.
FT DOMAIN 8363 8401 EGF-LIKE 151. 4049 4049 BY SIMILARITY.
FT DOMAIN 8402 8440 EGF-LIKE 152. 4061 4061 BY SIMILARITY.
FT DOMAIN 8441 8479 EGF-LIKE 153. 4073 4073 BY SIMILARITY.
FT DOMAIN 8480 8518 EGF-LIKE 154. 4085 4085 BY SIMILARITY.
FT DOMAIN 8519 8557 EGF-LIKE 155. 4097 4097 BY SIMILARITY.
FT DOMAIN 8558 8596 EGF-LIKE 156. 4109 4109 BY SIMILARITY.
FT DOMAIN 8597 8635 EGF-LIKE 157. 4121 4121 BY SIMILARITY.
FT DOMAIN 8636 8674 EGF-LIKE 158. 4133 4133 BY SIMILARITY.
FT DOMAIN 8675 8713 EGF-LIKE 159. 4145 4145 BY SIMILARITY.
FT DOMAIN 8714 8752 EGF-LIKE 160. 4157 4157 BY SIMILARITY.
FT DOMAIN 8753 8791 EGF-LIKE 161. 4169 4169 BY SIMILARITY.
FT DOMAIN 8792 8830 EGF-LIKE 162. 4181 4181 BY SIMILARITY.
FT DOMAIN 8831 8869 EGF-LIKE 163. 4193 4193 BY SIMILARITY.
FT DOMAIN 8870 8908 EGF-LIKE 164. 4205 4205 BY SIMILARITY.
FT DOMAIN 8909 8947 EGF-LIKE 165. 4217 4217 BY SIMILARITY.
FT DOMAIN 8948 8986 EGF-LIKE 166. 4229 4229 BY SIMILARITY.
FT DOMAIN 8987 9025 EGF-LIKE 167. 4241 4241 BY SIMILARITY.
FT DOMAIN 9026 9064 EGF-LIKE 168. 4253 4253 BY SIMILARITY.
FT DOMAIN 9065 9103 EGF-LIKE 169. 4265 4265 BY SIMILARITY.
FT DOMAIN 9104 9142 EGF-LIKE 170. 4277 4277 BY SIMILARITY.
FT DOMAIN 9143 9181 EGF-LIKE 171. 4289 4289 BY SIMILARITY.
FT DOMAIN 9182 9220 EGF-LIKE 172. 4301 4301 BY SIMILARITY.
FT DOMAIN 9221 9259 EGF-LIKE 173. 4313 4313 BY SIMILARITY.
FT DOMAIN 9260 9298 EGF-LIKE 174. 4325 4325 BY SIMILARITY.
FT DOMAIN 9299 9337 EGF-LIKE 175. 4337 4337 BY SIMILARITY.
FT DOMAIN 9338 9376 EGF-LIKE 176. 4349 4349 BY SIMILARITY.
FT DOMAIN 9377 9415 EGF-LIKE 177. 4361 4361 BY SIMILARITY.
FT DOMAIN 9416 9454 EGF-LIKE 178. 4373 4373 BY SIMILARITY.
FT DOMAIN 9455 9493 EGF-LIKE 179. 4385 4385 BY SIMILARITY.
FT DOMAIN 9494 9532 EGF-LIKE 180. 4397 4397 BY SIMILARITY.
FT DOMAIN 9533 9571 EGF-LIKE 181. 4409 4409 BY SIMILARITY.
FT DOMAIN 9572 9610 EGF-LIKE 182. 4421 4421 BY SIMILARITY.
FT DOMAIN 9611 9649 EGF-LIKE 183. 4433 4433 BY SIMILARITY.
FT DOMAIN 9650 9688 EGF-LIKE 184. 4445 4445 BY SIMILARITY.
FT DOMAIN 9689 9727 EGF-LIKE 185. 4457 4457 BY SIMILARITY.
FT DOMAIN 9728 9766 EGF-LIKE 186. 4469 4469 BY SIMILARITY.
FT DOMAIN 9767 9805 EGF-LIKE 187. 4481 4481 BY SIMILARITY.
FT DOMAIN 9806 9844 EGF-LIKE 188. 4493 4493 BY SIMILARITY.
FT DOMAIN 9845 9883 EGF-LIKE 189. 4505 4505 BY SIMILARITY.
FT DOMAIN 9884 9922 EGF-LIKE 190. 4517 4517 BY SIMILARITY.
FT DOMAIN 9923 9961 EGF-LIKE 191. 4529 4529 BY SIMILARITY.
FT DOMAIN 9962 10000 EGF-LIKE 192. 4541 4541 BY SIMILARITY.
FT DOMAIN 10001 10039 EGF-LIKE 193. 4553 4553 BY SIMILARITY.
FT DOMAIN 10040 10078 EGF-LIKE 194. 4565 4565 BY SIMILARITY.
FT DOMAIN 10079 10117 EGF-LIKE 195. 4577 4577 BY SIMILARITY.
FT DOMAIN 10118 10156 EGF-LIKE 196. 4589 4589 BY SIMILARITY.
FT DOMAIN 10157 10195 EGF-LIKE 197. 4601 4601 BY SIMILARITY.
FT DOMAIN 10196 10234 EGF-LIKE 198. 4613 4613 BY SIMILARITY.
FT DOMAIN 10235 10273 EGF-LIKE 199. 4625 4625 BY SIMILARITY.
FT DOMAIN 10274 10312 EGF-LIKE 200. 4637 4637 BY SIMILARITY.
FT DOMAIN 10313 10351 EGF-LIKE 201. 4649 4649 BY SIMILARITY.
FT DOMAIN 10352 10390 EGF-LIKE 202. 4661 4661 BY SIMILARITY.
FT DOMAIN 10391 10429 EGF-LIKE 203. 4673 4673 BY SIMILARITY.
FT DOMAIN 10430 10468 EGF-LIKE 204. 4685 4685 BY SIMILARITY.
FT DOMAIN 10469 10507 EGF-LIKE 205. 4697 4697 BY SIMILARITY.
FT DOMAIN 10508 10546 EGF-LIKE 206. 4709 4709 BY SIMILARITY.
FT DOMAIN 10547 10585 EGF-LIKE 207. 4721 4721 BY SIMILARITY.
FT DOMAIN 10586 10624 EGF-LIKE 208. 4733 4733 BY SIMILARITY.
FT DOMAIN 10625 10663 EGF-LIKE 209. 4745 4745 BY SIMILARITY.
FT DOMAIN 10664 10702 EGF-LIKE 210. 4757 4757 BY SIMILARITY.
FT DOMAIN 10703 10741 EGF-LIKE 211. 4769 4769 BY SIMILARITY.
FT DOMAIN 10742 10780 EGF-LIKE 212. 4781 4781 BY SIMILARITY.
FT DOMAIN 10781 10819 EGF-LIKE 213. 4793 4793 BY SIMILARITY.
FT DOMAIN 10820 10858 EGF-LIKE 214. 4805 4805 BY SIMILARITY.
FT DOMAIN 10859 10897 EGF-LIKE 215. 4817 4817 BY SIMILARITY.
FT DOMAIN 10898 10936 EGF-LIKE 216. 4829 4829 BY SIMILARITY.
FT DOMAIN 10937 10975 EGF-LIKE 217. 4841 4841 BY SIMILARITY.
FT DOMAIN 10976 11014 EGF-LIKE 218. 4853 4853 BY SIMILARITY.
FT DOMAIN 11015 11053 EGF-LIKE 219. 4865 4865 BY SIMILARITY.
FT DOMAIN 11054 11092 EGF-LIKE 220. 4877 4877 BY SIMILARITY.
FT DOMAIN 11093 11131 EGF-LIKE 221. 4889 4889 BY SIMILARITY.
FT DOMAIN 11132 11170 EGF-LIKE 222. 4901 4901 BY SIMILARITY.
FT DOMAIN 11171 11209 EGF-LIKE 223. 4913 4913 BY SIMILARITY.
FT DOMAIN 11210 11248 EGF-LIKE 224. 4925 4925 BY SIMILARITY.
FT DOMAIN 11249 11287 EGF-LIKE 225. 4937 4937 BY SIMILARITY.
FT DOMAIN 11288 11326 EGF-LIKE 226. 4949 4949 BY SIMILARITY.
FT DOMAIN 11327 11365 EGF-LIKE 227. 4961 4961 BY SIMILARITY.
FT DOMAIN 11366 11404 EGF-LIKE 228. 4973 4973 BY SIMILARITY.
FT DOMAIN 11405 11443 EGF-LIKE 229. 4985 4985 BY SIMILARITY.
FT DOMAIN 11444 11482 EGF-LIKE 230. 4997 4997 BY SIMILARITY.
FT DOMAIN 11483 11521 EGF-LIKE 231. 5009 5009 BY SIMILARITY.
FT DOMAIN 11522 11560 EGF-LIKE 232. 5021 5021 BY SIMILARITY.
FT DOMAIN 11561 11599 EGF-LIKE 233. 5033 5033 BY SIMILARITY.
FT DOMAIN 11600 11638 EGF-LIKE 234. 5045 5045 BY SIMILARITY.
FT DOMAIN 11639 11677 EGF-LIKE 235. 5057 5057 BY SIMILARITY.
FT DOMAIN 11678 11716 EGF-LIKE 236. 5069 5069 BY SIMILARITY.
FT DOMAIN 11717 11755 EGF-LIKE 237. 5081 5081 BY SIMILARITY.
FT DOMAIN 11756 11794 EGF-LIKE 238. 5093 5093 BY SIMILARITY.
FT DOMAIN 11795 11833 EGF-LIKE 239. 5105 5105 BY SIMILARITY.
FT DOMAIN 11834 11872 EGF-LIKE 240. 5117 5117 BY SIMILARITY.
FT DOMAIN 11873 11911 EGF-LIKE 241. 5129 5129 BY SIMILARITY.
FT DOMAIN 11912 11950 EGF-LIKE 242. 5141 5141 BY SIMILARITY.
FT DOMAIN 11951 11989 EGF-LIKE 243. 5153 5153 BY SIMILARITY.
FT DOMAIN 11990 12028 EGF-LIKE 244. 5165 5165 BY SIMILARITY.
FT DOMAIN 12029 12067 EGF-LIKE 245. 5177 5177 BY SIMILARITY.
FT DOMAIN 12068 12106 EGF-LIKE 246. 5189 5189 BY SIMILARITY.
FT DOMAIN 12107 12145 EGF-LIKE 247. 5201 5201 BY SIMILARITY.
FT DOMAIN 12146 12184 EGF-LIKE 248. 5213 5213 BY SIMILARITY.
FT DOMAIN 12185 12223 EGF-LIKE 249. 5225 5225 BY SIMILARITY.
FT DOMAIN 12224 12262 EGF-LIKE 250. 5237 5237 BY SIMILARITY.
FT DOMAIN 12263 12301 EGF-LIKE 251. 5249 5249 BY SIMILARITY.
FT DOMAIN 12302 12340 EGF-LIKE 252. 5261 5261 BY SIMILARITY.
FT DOMAIN 12341 12379 EGF-LIKE 253. 5273 5273 BY SIMILARITY.
FT DOMAIN 12380 12418 EGF-LIKE 254. 5285 5285 BY SIMILARITY.
FT DOMAIN 12419 12457 EGF-LIKE 255. 5297 5297 BY SIMILARITY.
FT DOMAIN 12458 12496 EGF-LIKE 256. 5309 5309 BY SIMILARITY.
FT DOMAIN 12497 12535 EGF-LIKE 257. 5321 5321 BY SIMILARITY.
FT DOMAIN 12536 12574 EGF-LIKE 258. 5333 5333 BY SIMILARITY.
FT DOMAIN 12575 12613 EGF-LIKE 259. 5345 5345 BY SIMILARITY.
FT DOMAIN 12614 12652 EGF-LIKE 260. 5357 5357 BY SIMILARITY.
FT DOMAIN 12653 12691 EGF-LIKE 261. 5369 5369 BY SIMILARITY.
FT DOMAIN 12692 12730 EGF-LIKE 262. 5381 5381 BY SIMILARITY.
FT DOMAIN 12731 12769 EGF-LIKE 263. 5393 5393 BY SIMILARITY.
FT DOMAIN 12770 12808 EGF-LIKE 264. 5405 5405 BY SIMILARITY.
FT DOMAIN 12809 12847 EGF-LIKE 265. 5417 5417 BY SIMILARITY.
FT DOMAIN 12848 12886 EGF-LIKE 266. 5429 5429 BY SIMILARITY.
FT DOMAIN 12887 12925 EGF-LIKE 267. 5441 5441 BY SIMILARITY.
FT DOMAIN 12926 12964 EGF-LIKE 268. 5453 5453 BY SIMILARITY.
FT DOMAIN 12965 13003 EGF-LIKE 269. 5465 5465 BY SIMILARITY.
FT DOMAIN 13004 13042 EGF-LIKE 270. 5477 5477 BY SIMILARITY.
FT DOMAIN 13043 13081 EGF-LIKE 271. 5489 5489 BY SIMILARITY.
FT DOMAIN 13082 13120 EGF-LIKE 272. 5501 5501 BY SIMILARITY.
FT DOMAIN 13121 13159 EGF-LIKE 273. 5513 5513 BY SIMILARITY.
FT DOMAIN 13160 13198 EGF-LIKE 274. 5525 5525 BY SIMILARITY.
FT DOMAIN 13199 13237 EGF-LIKE 275. 5537 5537 BY SIMILARITY.
FT DOMAIN 13238 13276 EGF-LIKE 276. 5549 5549 BY SIMILARITY.
FT DOMAIN 13277 13315 EGF-LIKE 277. 5561 5561 BY SIMILARITY.
FT DOMAIN 13316 13354 EGF-LIKE 278. 5573 5573 BY SIMILARITY.
FT DOMAIN 13355 13393 EGF-LIKE 279. 5585 5585 BY SIMILARITY.
FT DOMAIN 13394 13432 EGF-LIKE 280. 5597 5597 BY SIMILARITY.
FT DOMAIN 13433 13471 EGF-LIKE 281. 5609 5609 BY SIMILARITY.
FT DOMAIN 13472 13510 EGF-LIKE 282. 5621 5621 BY SIMILARITY.
FT DOMAIN 13511 13549 EGF-LIKE 283. 5633 5633 BY SIMILARITY.
FT DOMAIN 13550 13588 EGF-LIKE 284. 5645 5645 BY SIMILARITY.
FT DOMAIN 13589 13627 EGF-LIKE 285. 5657 5657 BY SIMILARITY.
FT DOMAIN 13628 13666 EGF-LIKE 286. 5669 5669 BY SIMILARITY.
FT DOMAIN 13667 13705 EGF-LIKE 287. 5681 5681 BY SIMILARITY.
FT DOMAIN 13706 13744 EGF-LIKE 288. 5693 5693 BY SIMILARITY.
FT DOMAIN 13745 13783 EGF-LIKE 289. 5705 5705 BY SIMILARITY.
FT DOMAIN 13784 13822 EGF-LIKE 290. 5717 5717 BY SIMILARITY.
FT DOMAIN 13823 13861 EGF-LIKE 291. 5729 5729 BY SIMILARITY.
FT DOMAIN 13862 13900 EGF-LIKE 292. 5741 5741 BY SIMILARITY.
FT DOMAIN 13901 13939 EGF-LIKE 293. 5753 5753 BY SIMILARITY.
FT DOMAIN 13940 13978 EGF-LIKE 294. 5765 5765 BY SIMILARITY.
FT DOMAIN 13979 14017 EGF-LIKE 295. 5777 5777 BY SIMILARITY.
FT DOMAIN 14018 14056 EGF-LIKE 296. 5789 5789 BY SIMILARITY.
FT DOMAIN 14057 14095 EGF-LIKE 297. 5801 5801 BY SIMILARITY.
FT DOMAIN 14096 14134 EGF-LIKE 298. 5813 5813 BY SIMILARITY.
FT DOMAIN 14135 14173 EGF-LIKE 299. 5825 5825 BY SIMILARITY.
FT DOMAIN 14174 14212 EGF-LIKE 300. 5837 5837 BY SIMILARITY.
FT DOMAIN 14213 14251 EGF-LIKE 301. 5849 5849 BY SIMILARITY.
FT DOMAIN 14252 14290 EGF-LIKE 302. 5861 5861 BY SIMILARITY.
FT DOMAIN 14291 14329 EGF-LIKE 303. 5873 5873 BY SIMILARITY.
FT DOMAIN 14330 14368 EGF-LIKE 304. 5885 5885 BY SIMILARITY.
FT DOMAIN 14369 14407 EGF-LIKE 305. 5897 5897 BY SIMILARITY.
FT DOMAIN 14408 14446 EGF-LIKE 306. 5909 5909 BY SIMILARITY.
FT DOMAIN 14447 14485 EGF-LIKE 307. 5921 5921 BY SIMILARITY.
FT DOMAIN 14486 14524 EGF-LIKE 308. 5933 5933 BY SIMILARITY.
FT DOMAIN 14525 14563 EGF-LIKE 309. 5945 5945 BY SIMILARITY.
FT DOMAIN 14564 14602 EGF-LIKE 310. 5957 5957 BY SIMILARITY.
FT DOMAIN 14603 14641 EGF-LIKE 311. 5969 5969 BY SIMILARITY.
FT DOMAIN 14642 14680 EGF-LIKE 312. 5981 5981 BY SIMILARITY.
FT DOMAIN 14681 14719 EGF-LIKE 313. 5993 5993 BY SIMILARITY.
FT DOMAIN 14720 14758 EGF-LIKE 314. 6005 6005 BY SIMILARITY.
FT DOMAIN 14759 14797 EGF-LIKE 315. 6017 6017 BY SIMILARITY.
FT DOMAIN 14798 14836 EGF-LIKE 316. 6029 6029 BY SIMILARITY.
FT DOMAIN 14837 14875 EGF-LIKE 317. 60
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 45 Seconds
(without alignments)
1801.960 Million cell updates/sec

Title: US-10-030-529A-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFTYNGMSYGASGVGEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organellae.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1342	100.0	257	16 Q9K2H6	Q9K2H6 haemophilus
2	1299.5	96.8	264	2 Q9K3B8	Q9K3B8 haemophilus
3	1289	96.1	271	2 Q9L916	Q9L916 haemophilus
4	1283	95.6	273	2 Q9L917	Q9L917 haemophilus
5	1262	94.0	248	2 Q9L919	Q9L919 haemophilus
6	1205	89.8	263	2 Q9L918	Q9L918 haemophilus
7	225	16.8	295	2 Q8RR69	Q8RR69 actinobacil
8	225	16.8	295	2 Q8RQ63	Q8RQ63 actinobacil
9	222	16.5	504	2 Q9LA56	Q9LA56 escherichia
10	222	16.5	511	9 Q9MC18	Q9MC18 bacterioph
11	220	16.4	392	2 Q9LA60	Q9LA60 escherichia
12	217	16.2	355	2 Q8KU1	Q8KU1 haemophilus
13	217	16.2	362	2 Q8KH7	Q8KH7 neisseria m
14	217	16.2	364	16 Q9JXK7	Q9JXK7 neisseria m
15	217	16.2	459	2 Q8VW24	Q8VW24 escherichia
16	209	15.6	391	2 Q8KU6	Q8KU6 neisseria m

17	209	15.6	398	2	Q8KI42	Q8KI42 neisseria m
18	208.5	15.5	487	2	Q9LA53	Q9LA53 escherichia
19	204	15.2	405	2	Q8KHP5	Q8KHP5 neisseria m
20	204	15.2	405	2	Q8KH85	Q8KH85 neisseria m
21	198	14.8	668	2	Q8GH86	Q8GH86 moraxella c
22	196	14.6	630	2	Q9XD55	Q9XD55 moraxella c
23	193	14.4	613	2	Q9XD53	Q9XD53 moraxella c
24	193	14.4	616	2	Q848S2	Q848S2 moraxella c
25	193	14.4	877	2	Q848S1	Q848S1 moraxella c
26	193	14.4	889	2	Q9L961	Q9L961 moraxella c
27	193	14.4	894	2	Q9L962	Q9L962 moraxella c
28	190.5	14.2	576	2	Q54407	Q54407 moraxella c
29	190.5	14.2	674	2	Q9XD51	Q9XD51 moraxella c
30	190.5	14.2	684	2	Q9L963	Q9L963 moraxella c
31	190.5	14.2	686	2	Q8RTB2	Q8RTB2 moraxella c
32	186	13.9	422	2	Q93KN4	Q93KN4 yersinia en
33	186	13.9	422	2	Q56930	Q56930 yersinia en
34	180.5	13.5	422	2	Q84GR6	Q84GR6 yersinia en
35	161	12.0	454	2	O85267	O85267 yersinia en
36	159.5	11.9	2712	16	Q9F3X5	Q9F3X5 pasteurella
37	132	9.8	155	16	Q8YJ81	Q8YJ81 brucella me
38	130	9.7	278	16	Q8FYM1	Q8FYM1 brucella su
39	125.5	9.4	144	16	Q8CL85	Q8CL85 yersinia pe
40	124.5	9.3	1299	16	Q9F3X6	Q9F3X6 pasteurella
41	119.5	8.9	641	16	Q8CKM1	Q8CKM1 yersinia pe
42	119.5	8.9	658	16	Q8ZH70	Q8ZH70 yersinia pe
43	116	8.6	516	2	Q93F81	Q93F81 escherichia
44	112.5	8.4	364	16	Q8ZGB6	Q8ZGB6 yersinia pe
45	110	8.2	873	2	Q9L960	Q9L960 moraxella c

ALIGNMENTS

RESULT 1

Q9K2H6 PRELIMINARY; PRT; 257 AA.
ID Q9K2H6
AC Q9K2H6
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Serum resistance protein DsrA (Serum resistance protein DsrA).
GN DSR A OR HD0769.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP542(CDC), 35000, and 406;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA."
RL Infect. Immun. 68:1608-1619(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF187001; AAF37813.1; -
DR EMBL; AF187001; AAF37807.1; -
DR EMBL; AF187005; AAF37811.1; -
DR EMBL; AF187005; AAF37811.1; -
DR EMBL; AF187005; AAF37811.1; -
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28524 MW; F497BF2CD5666938 CRC64;

Query Match 100.0%; Score 1342; DB 16; Length 257;
Best Local Similarity 100.0%; Pred. No. 5.4e-95;

```
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 180
Db 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 180
Qy 181 KELOTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVSRITDFTAKAGVAF 240
Db 181 KELOTLGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVSRITDFTAKAGVAF 240
Qy 241 NTYNGGMSYGASVGVEF 257
Db 241 NTYNGGMSYGASVGVEF 257

RESULT 2
Q9K3B8
ID Q9K3B8 PRELIMINARY; PRT; 264 AA.
AC Q9K3B8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIPA77; PubMed=10678980;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187003; AAF37809.1; -.
DR EMBL; AF187002; AAF37808.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 264 AA; 29444 MW; 11AF6C124D7ED126 CRC64;

Query Match 96.8%; Score 1299.5; DB 2; Length 264;
Best Local Similarity 95.8%; Pred. No. 1e-91;
Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 173
Db 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 180
Qy 174 HNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVSRITDFT 233
Db 181 HNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVSRITDFT 240
Qy 234 AKAGVAFNTYNGGMSYGASVGVEF 257
Db 241 AKAGVAFNTYNGGMSYGASVGVEF 264
```

RESULT 3

```
Q9L916
ID Q9L916 PRELIMINARY; PRT; 271 AA.
AC Q9L916;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1157;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187009; AAF37815.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 271 AA; 30091 MW; 5A9DC55C6DA6EF1C CRC64;
```

Query Match 96.1%; Score 1289; DB 2; Length 271;
Best Local Similarity 93.0%; Pred. No. 6.7e-91;
Matches 252; Conservative 0; Mismatches 5; Indels 14; Gaps 1;

```
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 171
Db 121 LYSYFNDLRHDFPKLVLDARISKKNQIDTTSKYLLELGTVDLSYRMEQNTNHNKNT 180
Qy 172 -----NTHNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGS 226
Db 181 HNINKNTHNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGS 240
Qy 227 RITDRFTAKAGVAFNTYNGGMSYGASVGVEF 257
Db 241 RITDRFTAKAGVAFNTYNGGMSYGASVGVEF 271
```

RESULT 4

```
Q9L917
ID Q9L917 PRELIMINARY; PRT; 273 AA.
AC Q9L917;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
```

```
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187008; AAF37814.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 273 AA; 30281 MW; 8C620D6B5544E3EE CRC64;

Query Match 95.6%; Score 1283; DB 2; Length 273;
Best Local Similarity 92.7%; Pred. No. 1.9e-90;
Matches 253; Conservative 0; Mismatches 4; Indels 16; Gaps 2;

Qy 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Qy 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPS--PILLYPMSDDPOLGINSROOLK 118
Db 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSISILLYPMSDDPOLGINSROOLK 120
Qy 119 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQ----- 171
Db 121 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINK 180
Qy 172 -----NTNINKLSKELQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGV 224
Db 181 NTNINKNTNINKLSKELQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGV 240
Qy 225 GSRITDRFTAKAGVAFTNYNGHSGYGVYEF 257
Db 241 GSRITDRFTAKAGVAFTNYNGHSGYGVYEF 273

RESULT 5
Q9L919 PRELIMINARY; PRT; 248 AA.
AC Q9L919;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DsrA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHIA;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187008; AAF37812.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
DR SEQUENCE 248 AA; 27815 MW; 245C809F24B8A815 CRC64;

Query Match 94.0%; Score 1262; DB 2; Length 248;
Best Local Similarity 99.6%; Pred. No. 7e-89;
Matches 242; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Qy 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 120
Db 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 120
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180
Db 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180

RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187004; AAF37810.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 248 AA; 27815 MW; 245C809F24B8A815 CRC64;

Query Match 94.0%; Score 1262; DB 2; Length 248;
Best Local Similarity 99.6%; Pred. No. 7e-89;
Matches 242; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Qy 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 120
Db 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 120
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180
Db 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180

Qy 181 KKLQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGVSGSRITDRFTAKAGVAF 240
Db 181 KKLQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGVSGSRITDRFTAKAGVAF 240
Qy 241 NTY 243
Db 241 NTF 243

RESULT 6
Q9L918 PRELIMINARY; PRT; 263 AA.
AC Q9L918;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DsrA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHIA;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187006; AAF37812.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
DR SEQUENCE 263 AA; 29170 MW; 858B91BF765B89F1 CRC64;

Query Match 89.8%; Score 1205; DB 2; Length 263;
Best Local Similarity 89.0%; Pred. No. 1.8e-84;
Matches 235; Conservative 8; Mismatches 13; Indels 8; Gaps 2;

Qy 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Db 1 MKIKCLVAVVGLACSTTTTMAQQPPKFAVSSLSYSDYDYGKGTWTSNNGGFDIKVPGI 60
Qy 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 120
Db 61 KMKPKWISKOATYLELQHYMPYTPVLVTPAGVSPSILLYPMSDDPOLGINSROOLK 119
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQ-----NT 173
Db 120 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKNT 179
Qy 174 HNINKLSKELQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGVSGSRITDRFT 233
Db 180 HNINKLSKELQTLGLANQSALSMVLPNGVGTSSAAVGGYRDKTALAIGVSGSRITDRFT 239
Qy 234 AKAGVAFNTYNGHSGYGVYEF 257
Db 240 AKAGVAFNTYNGHSGYGVYEF 263

RESULT 7
Q8RR69 PRELIMINARY; PRT; 295 AA.
AC Q8RR69;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein 100.
GN OMP100.
OS Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
```

```
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RA Komatsuzawa H., Asakawa R., Kawai T., Ochiai K., Fujiwara T.,
RT Taubman M.A., Kurihara H., Sugai M.;
RL "Identification of six major outer membrane proteins from
  Actinobacillus actinomycetemcomitans.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064943; BAB86905.1; -
DR InterPro; IPR008378; Ubiqu_surface.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
DR PRINTS; PR01804; UBIQUITOUISSP.
SQ SEQUENCE 295 AA; 32062 MW; C9A993C40D7E1C69 CRC64;

Query Match 16.8%; Score 225; DB 2; Length 295;
Best Local Similarity 31.8%; Pred. No. 3e-09;
Matches 55; Conservative 34; Mismatches 54; Indels 30; Gaps 2;

QY 111 GINRQQLKLNLYSYFNDLRHDPKLVLDARISKKNQID----- 149
DB 127 GNVRLVQNVDVRSSTENAARS-----RANEQKI AENKKA IENKADKADVEKNRADI AANSRA 182
QY 150 -----TISKYLLELGTYLDDSYRMEQNTNINKLSKELOTGLANQSA LMLVQPNVGK 204
DB 183 IATFRSSQNTIAALTTKVDRTNARTIDRLDSRVNLDKEVKNGLASQAALSGLFQPNVGS 242
QY 205 TSVSAAGVGYRDKTALAIGVGSRTTDRFTAKAGVAFNTYNGMSGYASGVGYEF 257
DB 243 LNLSSAAGVGYKSTALAVGSGYRNFQNVAAKAGVAVSTNGGSA TYNVGLNFEW 295

RESULT 8
QY Q8RQ63 PRELIMINARY; PRT; 295 AA.
AC Q8RQ63;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin/Invasin.
GN AA32-1-2.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OC actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RA Lepine G., Li L., Ellen R.P.;
RT "Cloning and characterization of three invasive genes of
  Actinobacillus actinomycetemcomitans.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316502; AAL91673.1; -
DR InterPro; IPR008378; Ubiqu_surface.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
DR PRINTS; PR01804; UBIQUITOUISSP.
SQ SEQUENCE 295 AA; 32034 MW; C9B75DD9DD7CCDBA CRC64;

Query Match 16.8%; Score 225; DB 2; Length 295;
Best Local Similarity 31.8%; Pred. No. 3e-09;
Matches 55; Conservative 34; Mismatches 54; Indels 30; Gaps 2;

QY 111 GINRQQLKLNLYSYFNDLRHDPKLVLDARISKKNQID----- 149
DB 127 GNVRLVQNVDVRSSTENAARS-----RANEQKI AENKKA IENKADKADVEKNRADI AANSRA 182
QY 150 -----TISKYLLELGTYLDDSYRMEQNTNINKLSKELOTGLANQSA LMLVQPNVGK 204
DB 183 IATFRSSQNTIAALTTKVDRTNARTIDRLDSRVNLDKEVKNGLASQAALSGLFQPNVGS 242
QY 205 TSVSAAGVGYRDKTALAIGVGSRTTDRFTAKAGVAFNTYNGMSGYASGVGYEF 257

us-10-030-529a-2.rspt
```

```
DB 243 LNLSSAAGVGYKSTALAVGSGYRNFQNVAAKAGVAVSTNGGSA TYNVGLNFEW 295

RESULT 9
QY Q9LA56 PRELIMINARY; PRT; 504 AA.
AC Q9LA56;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibC.
GN EIBC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
  activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151674; AAF63035.1; -
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05658; Hep_Hag; 4.
DR Pfam; PF03895; YadaA; 1.
SQ SEQUENCE 504 AA; 53158 MW; 6B5E192DD4B2771B CRC64;

Query Match 16.5%; Score 222; DB 2; Length 504;
Best Local Similarity 38.0%; Pred. No. 1e-08;
Matches 49; Conservative 26; Mismatches 52; Indels 2; Gaps 1;

QY 129 RHDFKLVLDARISKKNQIDITISKYLLELGTYLDDSYRMEQNTNINKLSKELOTGLA 188
DB 378 RIDYAVGVIDGRVTRNTQSIKNSKAIAANTRTIQQHSARLDSQQRQINENHKMRAAA 437
QY 189 NQSALSMVLQPNVGKTSVSAAGVGYRDKTALAIGVGSRTTDRFTAKAGVAFNTYNGMS 248
DB 438 QSAALTGLFQPSYGVGKFNATAAGVGYDQQAALAVGVGYRFEQTAAKAGAVFS--DGDAS 495
QY 249 YGASVGYEF 257
DB 496 WNVGVNFEF 504

RESULT 10
QY Q9MC18 PRELIMINARY; PRT; 511 AA.
AC Q9MC18;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibD.
GN EibD.
OS Bacteriophage P-EibD.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=120163;
RN [1]
RP SEQUENCE FROM N.A.
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
  activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151675; AAF63040.1; -
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05658; Hep_Hag; 3.
DR Pfam; PF03895; YadaA; 1.
```

RESULT 12	:
Q8KUJ1	:
ID Q8KUJ1	: PRELIMINARY; :
AC Q8KUJ1;	PRT; 355 AA.
DT 01-OCT-2002	(TEMBlrel. 22, Created)
DT 01-OCT-2002	(TEMBlrel. 22, Last sequence update)
DT 01-MAR-2003	(TEMBlrel. 23, Last annotation update)
DE Putative adhesin/invasin.	:
OS Neisseria meningitidis.	:
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;	:
OC Neisseriaceae; Neisseria.	:

```

Query Match      16.2%; Score 217; DB 2; Length 362;
-Best Local Similarity 39.4%; Pred. No. 1.6e-08;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;

Qy 127 DLRHDFKFLKVLDRARISKNQKIDITISKYLLLEGLTGILYDSDYRMMEQNTHNINKLSKEIQTG 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 DIKADIATNKAD--IAKNSARIDSLDK-----NVANLRKETQG 290

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: May 13, 2004, 08:28:38 ; Search time 61 Seconds
(without alignments)
10656.045 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 2048
Sequence: 1 ataaatcgtcattgacatt.....aagccgttaagtgaggac 1168

Scoring table:

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 2291136

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10030529/runat.13052004.081454.2741/app.query.fasta.1.1351
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPC1=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10030529 @CGN.1.1.13 @runat.13052004.081454.2741
-NCPU=6 -ICPU=3 -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	193	9.4	610	10	US-09-952-267-11	Sequence 11, Appl
2	193	9.4	624	10	US-09-952-267-7	Sequence 7, Appl
3	193	9.4	889	10	US-09-952-267-15	Sequence 15, Appl
4	190.5	9.3	573	10	US-09-952-267-3	Sequence 3, Appl
5	159.5	7.8	2712	12	US-10-282-122A-67070	Sequence 67070, A
6	115	5.6	2042	14	US-10-193-584-6	Sequence 6, Appl
7	114	5.6	2039	14	US-10-193-584-7	Sequence 7, Appl
8	110	5.4	873	10	US-09-952-267-13	Sequence 13, Appl
9	108	5.3	852	12	US-10-282-122A-62892	Sequence 62892, A
10	108	5.3	892	10	US-09-952-267-5	Sequence 5, Appl
11	107	5.2	831	10	US-09-952-267-1	Sequence 1, Appl
12	107	5.2	941	10	US-09-952-267-9	Sequence 9, Appl
13	104.5	5.1	496	9	US-09-738-626-5483	Sequence 5483, Ap
14	100.5	4.9	1098	9	US-09-797-862-32	Sequence 32, Appl
15	100.5	4.9	2353	9	US-09-797-862-33	Sequence 33, Appl
16	99	4.8	795	12	US-10-282-122A-63527	Sequence 63527, A
17	92.5	4.5	767	9	US-09-815-242-5899	Sequence 5899, Ap
18	92.5	4.5	767	9	US-09-815-242-13140	Sequence 13140, A
19	92	4.5	335	12	US-10-282-122A-60846	Sequence 60846, A
20	90.5	4.4	723	15	US-10-369-493-10942	Sequence 10942, A
21	90.5	4.4	819	12	US-10-282-122A-54896	Sequence 54896, A
22	90.5	4.4	827	12	US-10-282-122A-60889	Sequence 60889, A
23	90.5	4.4	1055	14	US-10-032-585-7652	Sequence 7652, Ap
24	90.5	4.4	1225	14	US-10-233-553-20	Sequence 20, Appl
25	90.5	4.4	3073	12	US-10-282-122A-49147	Sequence 49147, A
26	90	4.4	478	9	US-09-815-242-11674	Sequence 11674, A
27	90	4.4	512	12	US-10-282-122A-59775	Sequence 59775, A
28	90	4.4	1160	12	US-10-147-493-234	Sequence 234, App
29	90	4.4	1160	12	US-10-145-127-234	Sequence 234, App
30	90	4.4	1160	12	US-10-160-503-234	Sequence 234, App
31	90	4.4	1160	12	US-10-143-118-234	Sequence 234, App
32	90	4.4	1160	12	US-10-144-993-234	Sequence 234, App
33	90	4.4	1160	12	US-10-158-787-234	Sequence 234, App
34	90	4.4	1160	12	US-10-140-024-234	Sequence 234, App
35	90	4.4	1160	12	US-10-140-808-234	Sequence 234, App
36	90	4.4	1160	12	US-10-152-405-234	Sequence 234, App
37	90	4.4	1160	12	US-10-127-852A-234	Sequence 234, App
38	90	4.4	1160	12	US-10-127-900A-234	Sequence 234, App
39	90	4.4	1160	12	US-10-128-685A-234	Sequence 234, App
40	90	4.4	1160	12	US-10-131-820A-234	Sequence 234, App
41	90	4.4	1160	12	US-10-142-886-234	Sequence 234, App
42	90	4.4	1160	12	US-10-146-728-234	Sequence 234, App
43	90	4.4	1160	12	US-10-146-785-234	Sequence 234, App
44	90	4.4	1160	12	US-10-147-499-234	Sequence 234, App
45	90	4.4	1160	12	US-10-157-798-234	Sequence 234, App

ALIGNMENTS

RESULT 1
US-09-952-267-11
; Sequence 11, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 09/336,447
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-11

Alignment Scores:
Pred. No.: 2.72e-11 Length: 610
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-11 (1-610)

Qy 476 AACGATTAAAGACACGATTTTAAATTAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaLleThrLysAsnGly 504
Qy 536 CAAAATATTCATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db AsnAlaLleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 524
Qy 593 TCTTATCGTATGATGGAAACAAATACACAT-----AATATCAATAAG 634
Db AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 544
Qy 635 TTGCTAAAGAAATTCGCAAACTGGTTTAGCCCAACCAATCAGCATTTGCTATGTTAGTCAA 694
Db LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 564
Qy 695 CCAAATGGTGTAGCAAAACGAGCGTTTCTGCGGTAGGAGTTATAGAGATAAAACT 754
Db ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyGlyTyrGlySerLysSer 584
Qy 755 GCATTAGCCATTGTCGCTCAGCATCTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db AlaValAlaLleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 604
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTTATGCTGCTCTGTTGGTTATGAATTC 871
Db AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 624

RESULT 3
US-09-952-267-15
; Sequence 15, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-15

Alignment Scores:
Pred. No.: 3.01e-11 Length: 889
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-15 (1-889)

Qy 476 AACGATTAAAGACACGATTTTAAATTAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaLleThrLysAsnGly 769
Qy 536 CAAAATATTCATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db AsnAlaLleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 789
Qy 593 TCTTATCGTATGATGGAAACAAATACACAT-----AATATCAATAAG 634

Alignment Scores:
Pred. No.: 2.74e-11 Length: 624
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-7 (1-624)

Qy 476 AACGATTAAAGACACGATTTTAAATTAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaLleThrLysAsnGly 490
Qy 536 CAAAATATTCATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db AsnAlaLleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 510
Qy 593 TCTTATCGTATGATGGAAACAAATACACAT-----AATATCAATAAG 634
Db AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 530
Qy 635 TTGCTAAAGAAATTCGCAAACTGGTTTAGCCCAACCAATCAGCATTTGCTATGTTAGTCAA 694
Db LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 550
Qy 695 CCAAATGGTGTAGCAAAACGAGCGTTTCTGCGGTAGGAGTTATAGAGATAAAACT 754
Db ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyGlyTyrGlySerLysSer 570
Qy 755 GCATTAGCCATTGTCGCTCAGCATCTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db AlaValAlaLleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 590
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTTATGCTGCTCTGTTGGTTATGAATTC 871
Db AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 610

RESULT 2
US-09-952-267-7
; Sequence 7, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-7

Db 790 AspGlyArgValThrAlaLeuAsePThrLysValAsnAlaPheAspGlyArgGileThrAla 809
|||
|||
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTAGTCCAA 694
|||
|||
Db 810 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 829
|||
|||
Qy 695 CCAATGGTGTAGCAGAAACGAGGTTTCTGCTGCGGTAGGAGGTATACAGATAAACT 754
|||
|||
Db 830 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyGlyTyrGlySerLysSer 849
|||
|||
Qy 755 GCATTAGCCATTGGTGTGCTCACGCACTACTGATCGCTTTACCGCTAAAGCGGTGTA 814
|||
|||
Db 850 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 869
|||
|||
Qy 815 CGGTTCAATACCTAC---AATGGCGGATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
|||
|||
Db 870 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 889
|||
|||
RESULT 4
US-09-952-267-3
; Sequence 3, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ASBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-3
Alignment Scores:
Pred. No.: 4,93e-11 Length: 573
Score: 190.50 Matches: 52
Percent Similarity: 47.62% Conservative: 18
Best Local Similarity: 35.37% Mismatches: 62
Query Match: 9.30% Indels: 15
DB: 10 Gaps: 3
US-10-030-529A-1 (1-1168) x US-09-952-267-3 (1-573)
Qy 476 AACGATTAGACAGCATTTTAATTAATAAGTCTTCATGCACGTAATTCACAAATAAA 535
|||
|||
Db 427 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 446
|||
|||
Qy 536 CAAAATATGATACATAAGTAATAATATTACTGAACCTGGT----- 577
|||
|||
Db 447 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 466
|||
|||
Qy 578 -----ACTTATTAGATGATCTTATCGTATGATGAACAAATAACACAT--- 622
|||
|||
Db 467 AspGlyArgValThrAlaLeuAsePThrLysValAsnAlaLeuAspThrLysValAsnAla 486
|||
|||
Qy 623 -----AATCAATAAGTCTTAAGAATGCAAACTGGTTAGCCAACTCA 673
|||
|||
Db 487 PheAspGlyArgIleThrAlaLeuAspSerLysValGluAsnGlyMetAlaAlaGlnAla 506
|||
|||
Qy 674 GCATTGTCTATTAGTGCAACCAATGGTGGCAAAACGAGCGTTCTGCTGCGGTA 733
|||
|||
Db 507 AlaLeuSerGlyLeuPheGlnProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeu 526
|||
|||

Qy 734 GGAGGTTATAGATAAACTGCATTAGCACTGGTGTGGCTCAGCATTAATCTGATCGC 793
|||
|||
Db 527 GlyGlyTyrGlySerLysSerAlaValAlaIleGlyAlaGlyTyrArgValAsnProAsn 546
|||
|||
Qy 794 TTTCACCGCTAAACGCGGTGTAGCGTTCAATACCTAC---AATGGCGCATGCTCTTATGGT 850
|||
|||
Db 547 LeuAlaPheLysAlaGlyAlaAlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsn 566
|||
|||
Qy 851 GCTTCCTGTTGGTTATGAATTC 871
|||
|||
Db 567 IleGlyValAsnTyrGluPhe 573
|||
|||

RESULT 5

US-10-282-122A-67070
; Sequence 67070, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67070
; LENGTH: 2712
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67070

Alignment Scores:
Pred. No.: 1,44e-07 Length: 2712
Score: 159.50 Matches: 45
Percent Similarity: 51.52% Conservative: 40
Best Local Similarity: 27.27% Mismatches: 49
Query Match: 7.79% Indels: 31
DB: 12 Gaps: 6

US-10-030-529A-1 (1-1168) x US-10-282-122A-67070 (1-2712)

```
QY 422 GATCAACTTTGGAATAAT-----CGGCAGCAGCTGAAATTTGAAT 460
Db 2564 AspVallyGlyIleAsnValThrGluAsnLeuThrVallyGluGlyAlaIlyIleAsn 2583
QY 461 TTGTATAGTTATTTAAACAT-----TTAAGACACGATTTTAAATA 502
Db 2584 MetGlyAsnAsnValIleAspGlyValAlaAspGlyGluValAsnAlaThrSerlyGln 2603
QY 503 AAAGTCTTGTGACGCTATTTCCAAAATAAACAATAATTTAGTACTATAGTAAATAT 562
Db 2604 AlaValAsnGlySerGlnLeuHisValGlnGlnValAsnAsnGlnAlaThrAla 2623
QY 563 TTACTAGAACTGGTACTTTATTTAGATGATTTCTTATCGTATGATGGAACAAAATACAT 622
Db 2624 IleAsnLysLeuGlyAspHis-----2630
QY 623 ATATCAATTAAGTTGCTTAAAGAAATGCAACTGTTTAGCCCAACCAATCAGCATTTGCT 682
Db 2631 ---IleAsnLysValAspLysAspLeuArgAlaGlyIleAlaGlyAlaThrAlaValAla 2649
QY 683 ATGTTAGTCAACCAATCGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTAT 742
Db 2650 PheLeuGlnArgProGlnGluAlaGlyLysSerIleValSerLeuGlyValGlySerTyr 2669
QY 743 AGAGATAAACTGCAATGACCAATGTTGTCGCTCAGCATTTACTGAT-----CGCTTT 796
Db 2670 ArgSerGluSerAlaIleAlaValGlyTyr---AlaArgAsnSerAspAsnAsnLysIle 2688
QY 797 ACGGCTAAAGCGGTTAGCGTTCAATCACTACATGCGCGCATGCTTTATGCGTCTCT 856
Db 2689 SerIleLysLeuGlyGlyGlyMetAsnSer---ArgGlyAspValAsnPheGlyGlySer 2707
QY 857 GTTGGTTATGAATTC 871
Db 2708 IleGlyTyrGlnTrp 2712

RESULT 6
US-10-192-584-6
; Sequence 6, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiichi
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukuaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
```

```
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 2042 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-192-584-6

Alignment Scores:
Pred. No.: 0.00702 Length: 2042
Score: 115.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.62% Indels: 6
DB: 14 Gaps: 3

US-10-030-529A-1 (1-1168) x US-10-192-584-6 (1-2042)
QY 569 GAACCTGGTACTTATTTAGATGATTTCTTATCGTATGATGGAACAAATACACATAATATC 628
Db 1942 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1961
QY 629 AATAAGTTGCTTAAAGAAATGCAACTGTTTAGCCCAACCAATCAGCATTTCTATGTTA 688
Db 1962 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerIleu 1981
QY 689 GTGCAACCAATGCTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGAT 748
Db 1982 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 2001
QY 749 AAAAATGCTAGCATTTGTTGTCGCTCAGCATTTACTGATCGCTTTTACCGCTAAAGCG 808
Db 2002 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2018
QY 809 GGTGTA-----GCGTTCAATACCTACATGCGCGCATGCTTATGGTGTCTCTGTT 859
Db 2019 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2038
QY 860 GGTATGAATTC 871
Db 2039 GlyPheSerPhe 2042

RESULT 7
US-10-192-584-7
; Sequence 7, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiichi
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukuaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
```

SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/192,584
FILING DATE: 11-Jul-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/077,098
FILING DATE: 19-May-1998
APPLICATION NUMBER: PCT/JP97/03222
FILING DATE: 12-SEP-1997
APPLICATION NUMBER: JP 27,148/1996
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: KORNEAU, Anne M.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TOKUNAGA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2039 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-192-584-7

Alignment Scores:
Pred. No.: 0.00895 Length: 2039
Score: 114.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.57% Indels: 6
DB: 14 Gaps: 3

US-10-030-529A-1 (1-1168) x US-10-192-584-7 (1-2039)

QY 569 GAACGGTACTTATTAGATGATCTTATCGTATGATGAAACAAATACACATATATC 628
DB 1939 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgile 1958
QY 629 AATAAGTTGCTAAGAAATGCAACTGTTTAGCCAAACCAATCAGCATTTGCTATGTTA 688
DB 1959 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1978
QY 689 GTCAACCAATGGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGAT 748
DB 1979 GlnAsnValAlaLeuProGlyLysThrThrIleSerValGlyThrAlaThrPheLysGly 1998
QY 749 AAACGTGATTAGCATTTGTGCGCTACGCAATTACTGATCGCTTTACCGCTAAAGCG 808
DB 1999 GlnAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2015
QY 809 GGTGTA-----GGTTCAATACCTACAAATGCGCGCATGCTTATGTTGCTGCTGTT 859
DB 2016 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2035
QY 860 GGTATGAAATTC 871
DB 2036 GlyPheThrPhe 2039

RESULT 8

US-09-952-267-13
; Sequence 13, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.

FILE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY.024
CURRENT APPLICATION NUMBER: US/09/952,267
CURRENT FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: 09/336,447
PRIOR FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 873
TYPE: PRT
ORGANISM: Moraxella catarrhalis
US-09-952-267-13

Alignment Scores:
Pred. No.: 0.019 Length: 873
Score: 110.00 Matches: 42
Percent Similarity: 47.37% Conservative: 30
Best Local Similarity: 27.63% Mismatches: 52
Query Match: 5.37% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-13 (1-873)

QY 428 CTTGGAATAAATCGCGAGCAGCTGAAATTCATATGATGATTTTAAACGATTTAAGA 487
DB 745 IleAlaThrAsnLysGlnGluLeu-----IleLeuGln 755
QY 488 CACGATTTTAAATTAAGTTCTTGTATGACGATTTTCCAAAATAAACAATATTGAT 547
DB 756 HisAsp---ArgLeuAsnArgileAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 774
QY 548 ACTATAAGTAATATTTACTAGNACTGGGTACTTATTAGATGATTTCTTATCGTATGATG 607
DB 775 -----GlnLeuGly-----TyrAlaLeuLys 781
QY 608 GAACAAATACACATAATATCAATAGTTGTCTAAA---GAATTCAAACT-----GGT 658
DB 782 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 801
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAACCAATGGTGTAGCAAAACGAGC 718
DB 802 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 821
QY 719 GTTCTGCTGCGGTAGGAGGTTATAGAGATAAACTGCAATGCCATTGGT---GTCCGC 775
DB 822 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaValSerLeuGlyAlaAlaGly 841
QY 776 TCACGCAATTACTCATCGCTTTACCGCTAAAGCGGGGTAGCGTTCAATACCTACAATGGC 835
DB 842 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 860
QY 836 GGCATGCTTATGCTGCTTCTGTTGTTGTTATGAATTC 871
DB 861 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 872

RESULT 9

US-10-282-122A-62892
; Sequence 62892, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62892
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62892

Alignment Scores:
Pred. No.: 0.0308 Length: 852
Score: 108.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.27% Indels: 18
DB: 12 Gaps: 8

US-10-030-529A-1 (1-1168) x US-10-282-122A-62892 (1-852)
QY 437 AATCGGCAGCAGCTGAATTTGTTATAGTTATTTTAAACGATTTAACAGCAGATTTT 496
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 496
QY 497 AAATTAAGTCTGTGATGTCAGCTATTTCACAAATTAACAAATATTTT 544
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 544
QY 716 GluileGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuLysGlnHis 735
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 735
QY 545 GATACTATAAGTAATTTTACTAGAACTGGTACTTATTAGATGAT 592
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 592
QY 736 AspArgLeuAsnArg---IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 754
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 754
QY 593 ---TCTTATCGTATGATGGAACAAATACACATAATATCAATAGTTGTCTAAA---GAA 646
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 646
QY 755 LeuGlyTyrAlaLeuLysGluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 774
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 774
QY 647 TTGCAAACT-----GGTTTAGCCCAACCAATCAGCATTTGTCTATTGTTAGTCACCAAT 700
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 700
QY 775 ArgGlnThrAlaGlyIleAlaAsnAlaIleAlaThrLeuProSerProSer 794
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 794
QY 701 GGTGTAGCAAAACGAGCGTTCTGCTCGCGTAGAGGTTATAGATATAAACTGCATTA 760
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 760
QY 795 ArgAlaGlyGluHisHisValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaVal 814
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 814
QY 761 GCCATTGCT---GTCGGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCG 817
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 817
QY 815 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSer 834
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 834
QY 818 TTCAATACCTCAATGCGCGCATGCTTTATGCTGCTTCTGTTGTTATGAATTC 871
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 871
```

```
Db 835 TrpSerAsp---AlaGlyGlyLeuSerGlyValGlyGlySerTyrArgTrp 851
RESULT 10
US-09-952-267-5
; Sequence 5, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCV.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-5

Alignment Scores:
Pred. No.: 0.0311 Length: 892
Score: 108.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.27% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-5 (1-892)
QY 428 CTTGGAATTAATCGGCAGCAGCTGAATTTGAATTTGTATAGTTATTTTAAACGATTTAAGA 487
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 487
QY 488 CACGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAAAATAACAAAAATATTGAT 547
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 547
QY 777 -----ArgLeuAsnGlnIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 793
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 793
QY 548 ACTATAAGTAATTTTACTAGAACTGGTACTTATTAGATGATCTTATCGTATGATG 607
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 607
QY 794 -----GlnLeuGly-----TyrAlaLeuLys 800
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 800
QY 608 GAACAAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCCAACT-----GGT 658
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 658
QY 801 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 820
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 820
QY 659 TTAGCCAAACCAATCAGCATTTGTCTATGTTAGTCACCAATGGTGTAGGCAAAACGAGC 718
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 718
QY 821 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 840
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 840
QY 719 GTTTCGTGCGGTAGGAGTTATAGATATAAACTGCATTTAGCCATTTGTT---GTCGCGC 775
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 775
QY 841 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaGly 860
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 860
QY 776 TCACGCATTACTGATCGCTTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACCAATGCC 835
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 835
QY 861 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 879
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 879
QY 836 GCGATGCTTTATGGTGTCTTCTGTTGGTTATGAATTC 871
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 871
QY 880 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 891
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 891
RESULT 11
US-09-952-267-1
; Sequence 1, Application US/09952267
; Publication No. US20030032772A1
```

```

; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-1

Alignment Scores:
Pred. No.: 0.039 Length: 831
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-1 (1-831)
QY 428 CTTGGAATAATCGGCAGCAGCTGAATTTGATAGTATTTTAAAGTTTAAAGA 487
Db 703 IleAlaThrAsnLysGlnGluLeuLeuGln-----AsnAsp----- 715
QY 488 CACGATTTAAATTAAGTTCTTGATGCACGTATTTCCAAATAAACAATAATATGAT 547
Db 716 -----ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 732
QY 548 ACTATAAGTAAATATTACTAGAACCTGGGTACTTATTAGATGATTTCTTCATGATG 607
Db 733 -----GlnLeuGly-----TyrAlaLeuLys 739
QY 608 GAACAAATACACATAATCAATAAGTTCTCTTAA-----CAATGCAAACT-----GGT 658
Db 740 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 759
QY 659 TTAGCCAAACCAATCAGCATTTCTTATGTAGTGCAACCAATGCTGTAGCAAAACGAGC 718
Db 760 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 779
QY 719 GTTCTGCTCGGTAGGAGGTTATAGAGATAAAACCTGCAATAGCCATTTGGT---GTCGGC 775
Db 780 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 799
QY 776 TCAGCATTAATCAATCGCTTACCGTAAAGCGGTGTCAGGTTCAATCACTCAATGCG 835
Db 800 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 818
QY 836 GGATGCTTTATGCTGCTCTGTTGTTATGAATTC 871
Db 819 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 830

RESULT 12
US-09-952-267-9
; Sequence 9, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.

; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-9

Alignment Scores:
Pred. No.: 0.0403 Length: 941
Score: 107.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.22% Indels: 18
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-9 (1-941)
QY 437 AATCGGCAGCAGCTGAATTTGATAGTATTTTAAACGATTTAAACACGATTTT 496
Db 788 AsnLysThrAlaIleGluGlnAsnIleAsn-----ArgThrValAlaAsnGlyPhe 804
QY 497 AAATTAATAAGTTCTTGATGCACGTATTTCCAAATAAACAATAATATT----- 544
Db 805 GluIleGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuLeuGlnAsn 824
QY 545 GATACATATAAGTAAATATTACTAGAACCTGGGTACTTATTAGATGAT----- 592
Db 825 AspArgLeuAsnGln---IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 843
QY 593 ---TCTTATCGTATGATGCAACAAATAACATAATATCAATATCAATAGTTGCTTAA--GAA 646
Db 844 LeuGlyTyrAlaLeuLysGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 863
QY 647 TTCCAAACT-----GGTTAGCAACCAATCAGCATTTGCTATGTAGTCAACCAAAAT 700
Db 864 ArgGlnThrAlaGlyGlyIleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSer 883
QY 701 GGTGTAGGCAAAACGACGCTTCTGTCGGGTAGGAGGTTATAGAGATAAAACCTGCAATTA 760
Db 884 ArgAlaGlyGluHisValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaVal 903
QY 761 GCATTTGGT---GTCGGCTCAGCATTAATGCTGCTTTACCGCTAAAGCGGTGTAGCG 817
Db 904 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSer 923
QY 818 TTCATTAACCTACAATGCGCATGCTCTTATGCTGCTTCTGTTGTTATGAATTC 871
Db 924 TrpSerAsp---AlaGlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 940

RESULT 13
US-09-738-626-5483
; Sequence 5483, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```



```
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5483
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5483

Alignment Scores:
Pred. No.: 0.0627 Length: 496
Score: 104.50 Matches: 67
Percent Similarity: 40.00% Conservative: 43
Best Local Similarity: 24.36% Mismatches: 104
Query Match: 5.10% Indels: 61
DB: 9 Gaps: 16

US-10-030-529A-1 (1-1168) x US-09-738-626-5483 (1-496)
Qy 173 CCAAGTTTGTGAGTATCT-----TCTTGTATAGCTATGATGAC 217
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 148 ProllysAlaAlaAspValSerGlnAlaThrValGlySerIleTyrSer----- 163
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 218 TATGTAAGGTAATAATGGACTTGTCTAATGAAGC-----GGTTTCGATATAAAGTG 271
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 164 PheGlyAspGlyArgIleGlyLeuMetLeuGluAlaProGluGlyPheAspMetGlnVal 183
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 272 CCAGGGATTAATAATGAAG---CCAAAAAGATGGATTTCTAAACAGGCTACTTATCTTGAA 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 184 LeuAsnIleThrMetArgAlaLysLysGluLeuValGluValAspSerThrAsnLeuThr 203
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 329 TTACAGCATATATGCCCTTATACCTCTCTCTCGTACATATGCTCTGGGTTCTCCT 388
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 204 SerGluAsnMetLys-----AlaGlnGlyTyrAspAsp 214
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 389 AGC-----CCTATAGTATTATCCGATGCTCGAT----- 418
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 215 SerPheIleAspHisAspGlnTrpArgIleMetLeuAsnProAlaAlaGluThrLeuAla 234
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 419 -----CCTGATCAACTTGAATAAATCGGCAGCAGCTGAAATTTGATAGTTAT 472
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 235 ThrLeuPro-----AlaLeuAsnAspGluGlnIleAspLeuArgAlaLeu---Tyr 250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 473 TTTAAGCATTTAAGCACAGCATTTAATAA---AAAGTTCTTGATGCAGGTATT 523
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 251 TrpGlnThrSerMetThrAspThrGluValGlyHisArgGlnAlaLeuAspThrSerLeu 270
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 524 TCCAAAAATAACAAAATATTGATGATAAGTAAATATTTTACTAGAACTGGT----- 577
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 271 SerArgMetAsnAlaLeuLeuGlyArgIleGluAsnTyrLysGluAlaLeuGlyValThr 290
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 578 -----ACTTATTTAGATGATCTTATCGTATGATGAACAAAATACACATATATCAAT 631
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 291 SerAlaThrIleMetGlyGluArgValSerLeuIleGluProIleProHis-----Glu 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 632 AAGTTGTCTAAGAAATGCAAACTGGTTTACCCCAACCAATCAGCATTTCTATGTTAGTG 691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 309 SerIleArgGlyAspArgIleGluGlyLeuAlaSerValAspValThrLysGlnLeuAla 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 692 CAA---CAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAT 748
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 329 GluAspProThrGly-----LysSerValProGlyTrpGluAsp 341
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 749 AAAACTGCATTAGCGATTTGGTGTGCGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCG 808
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

; 342 LysThrPheMetAlaAlaSerLeuAspAsnLeuMetSerGlu----- 355
Qy 809 GGTGTAGCGTTCAATACCTACAAATGCGCATGCTTATGCTGCTCTGTTGGTTATGAA 868
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 356 GlyTyrGlnTyrAsnSerTyrArgLeuValThrAspHisIlePheGlyGlyTyrAsp 375
Qy 869 TTCTAATCATTAACCTTAATCACTAATCGTTTGGTTTATATAATAA 913
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 376 Phe---AspTrpArgLeuGlyThrThrLeuAspGlyTyrLeuLys 389

RESULT 14
US-09-797-862-32
; Sequence 32, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOKON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-32

Alignment Scores:
Pred. No.: 0.205 Length: 1098
Score: 100.50 Matches: 50
Percent Similarity: 32.88% Conservative: 23
Best Local Similarity: 22.52% Mismatches: 90
Query Match: 4.91% Indels: 59
DB: 9 Gaps: 5

US-10-030-529A-1 (1-1168) x US-09-797-862-32 (1-1098)
Qy 227 GGTAAATGGAGTCTGCTAATGAAGCGGTTTCGATATATAAGTGCCAGGATTAAATG 286
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 929 GlyLysTrp-----TyrHisAlaLysAlaAspGlyThrAlaAsp 941
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 287 AAGCCAAAAGATGGATTTCTAAACAGGCTACTTATCTTGAATTCACGATTTATGCT 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 942 LysThrLysGlyGluValSerAsnAspLysValSerThrAspGluLysHisValValSer 961
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 347 TATACCTCTGTTCTCGTGACATATGCTCCGCGGCTTCTCTAGCCCTATATCTATAT 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 962 LeuAspProAsnAspGlnSerLysGlyValValLysAspAsnValAlaAsnGly 981
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 407 CCGATGTCGTGATCTGATCAACTTGGAAATAAATCGGCAGCAGCTG----- 451
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 982 AspIleSerAlaThrSerThrAspAlaIleAsnGlySerGlnLeuTyrAlaValAlaLys 1001
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 452 ---AAATTTGAATTTGTATAGTTATTTTAAACGATTTAAACACACGATTTTAAAGTT 508
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1002 GlyValThrAsnLeuAlaGlyGlnValAsnAsnLeuGlu----- 1014
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 509 CTTGATCGCATGATTTCCAAAAATAAACAATAATTTGATCTATATAAGTAAATATTACTA 568
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1014 ----- 1014
Qy 569 GAACCTGGTACTTATTTTAGATGATTTCTTATCGTATGATGGAACAAATACACATAATATC 628
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1015 -----GlyLysVal 1017
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 629 AATAAGTTGTCTAAGAAATGCAAACTGGTTTACGCCAACCAATCAGCATTTGCTATGTTA 688
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db	1018	AsnLysValGlyLysArgAlaAspAlaGlyThrAlaSerAlaLeuAlaAAsrGlnLeu	1037
Qy	689	GTGCAACCAANTGGTGTAGGCAAAACAGAGGGTTCTCTCGCGGTAGGAGGTTATATAGAGAT	748
Db	1038	ProGlnAlaThrMetProGlyLysSerMetValAlaIleAlaGlySerSerTyrGlnGly	1057
Qy	749	AAATCGATTCATAGGCATTGGTGTGCGCTCAGCATTTACTCATCGCTTTCACGCTAAAGCG	808
Db	1058	GlnAsnGlyLeuAlaIleGlyVal---SerArgIleSerAspAsnGlyLysValIleIle	1076
Qy	809	GGTGACCGTTCAATCACTACATGGCGCATGTCT---TATGGTGCTTCTGTGGTTAT	865
Db	1077	ArgLeuSerGlyThrThrAsnSerGlnGlyLysThrGlyValAlaAlaGlyValGlyTyr	1096
Qy	866	GAATTC	871
Db	1097	GlnTirp	1098

RESULT 15

```

US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PR1
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

```

Alignment Scores:

Pred. No.:	0.251	Length:	2353
Score:	100.50	Matches:	51
Percent Similarity:	33.78%	Conservative:	25
Best Local Similarity:	22.67%	Mismatches:	88
Query Match:	4.91%	Indels:	61
DB:	9	Gaps:	6

US-10-030-529A-1 (1-1168) x US-09-797-862-33 (1-2353)

Qy	221	GGTAAGGGTAAATGCGATCGTGTCTAAAT---GAAGGCGGTTTTCGATATTAAGAATGCCAGGG	217
Db	2182	GlyAlaAspLysTrpTyTyrThrAsnAlaAspGlyAlaAlaAsp-----	2196
Qy	278	ATTAATAATGAAGCAAAAGAAATGGATTTCTAAACAGCGTACTTATCTTGGATTTACAGCAT	337
Db	2197	-----LysThrLysGlyGluValSerAsnAspLysValSerThrAspGluLysHis	2213
Qy	338	TATATGCTTTATACTCTGTTCTCTGTCGACATATGCTCTCGCGGTTTCTCTAGTCCCTATA	397
Db	2214	ValValArgLeuAspProAsnAsnGlnSerAsnGlyLysGlyValValIleAspAsnVal	2233
Qy	398	CTGTTATATCCGATGCTCGATCTCTGATCAACTTGGATAAATCGGCAGCAGCTG-----	451
Db	2234	AlaAsnGlyGluLeuSerAlaThrSerThrAspAlaIleAsnGlySerGlnLeuTyrAla	2253
Qy	452	-----AAATTGAATTTGTATAGTATTTTAAACGATTTAAAGACACGATTTTAA	499
Db	2254	ValAlaLysGlyValThrAsnLeuAlaGlyGlnValAsnLeuGlu-----	2269

Qy	500	TTAAAAGTCTTGATGCACGTAATTTC	CAAAAAATAAACAAATAATTGATACTA	TAACTAA	559	
Dd	2269	-----	-----	-----	2269	
Qy	560	TATTTACTAGA	CTGGGTACTTATTTTAGATGATCTT	TATCGTATGATGGA	CAAATAACA	619
Dd	2269	-----	-----	-----	2269	
Qy	620	CATAATATCAATAAGTTGTCT	CTAAGAATTC	GCAAACTGGTTAGCCCAACCAATCAGCAT	TG	679
Dd	2270	GlyIysValAsnIysValGlyIys	ArgAlaSerPalaGlyThrAlaSerAlaLeuAlaAla			2289
Qy	680	TCATGTTAGTGCACCAATGTGT	AGGCAAAACGAGCGTTTTCTGCTGGGTAGGAGGT			739
Dd	2290	SerGlnLeuProGlnAlaThrMet	ProGlyLysSerMetValAlaIleAlaGlySerSer			2309
Qy	740	TATAGATATAAACTGCATTAGC	ATTAGCATTGGTTCGGCTCAGGCATTACTGATCGCTTACC			799
Dd	2310	TyrGlnGlyIleAsnGlyLeuAla	IleGlyVal---SerArgIleSerAspAsnGlyLys			2328
Qy	800	GCTAAAGCGGGTGTAGCGTTCA	ATCAACTCAAGTGGCGGATGTC---	TATGGTGCTTCT		856
Dd	2329	ValIleIleArgLeuSerGlyThr	ThrAsnSerGlnGlyIysThrGlyValAlaAlaGly			2348
Qy	857	GTGGTTATGAATTC				871
Dd	2349	ValGlyTyrGlnTrp				2353

Search completed: May 13, 2004, 08:40:37
Job time : 83 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:22:52 ; Search time 20 Seconds
(without alignments)
6029.913 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 2048
Sequence: 1 ataatacgtcattgacatt.....aagccgttaaggtcgagac 1168

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.P/US10030529/runat 13052004 081453 2656/app_query.fasta_1.1351
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529 @CGN 1.1 27 @runat 13052004 081453 2656 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGJOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	9.4	610	4	US-09-336-447A-11
2	193	9.4	624	4	US-09-336-447A-7
3	192	9.4	889	4	US-09-336-447A-15
4	192	9.4	867	4	US-09-540-236-2676
5	190.5	9.3	573	4	US-09-336-447A-3
6	134	6.7	436	4	US-09-489-037A-11785
7	123.5	6.1	568	4	US-09-543-681A-6966
8	115	5.6	2042	4	US-09-077-098A-6
9	114	5.6	2039	4	US-09-077-098A-7
10	110	5.4	873	4	US-09-336-447A-13
11	108	5.3	892	4	US-09-336-447A-5
12	107	5.2	831	4	US-09-336-447A-1

13	107	5.2	878	4	US-09-540-236-3401	Sequence 3401, Ap
14	107	5.2	941	4	US-09-336-447A-9	Sequence 9, Appli
15	104	5.1	1002	4	US-09-268-347-24	Sequence 24, Appli
16	104	5.1	1004	4	US-09-268-347-30	Sequence 30, Appli
17	103.5	5.1	616	4	US-09-268-347-38	Sequence 38, Appli
18	103	5.0	1104	4	US-09-268-347-28	Sequence 28, Appli
19	103	5.0	1104	4	US-09-268-347-34	Sequence 34, Appli
20	102	5.0	679	3	US-08-913-942-15	Sequence 15, Appli
21	102	5.0	679	4	US-09-268-347-26	Sequence 26, Appli
22	102	5.0	1094	4	US-09-268-347-32	Sequence 32, Appli
23	101	4.9	1690	4	US-09-595-684B-39	Sequence 39, Appli
24	100.5	4.9	1098	1	US-08-409-995-2	Sequence 2, Appli
25	100.5	4.9	1098	3	US-08-685-467-2	Sequence 2, Appli
26	100.5	4.9	1098	3	US-09-377-155-32	Sequence 32, Appli
27	100.5	4.9	1098	3	US-08-913-942-2	Sequence 2, Appli
28	100.5	4.9	1098	4	US-09-669-974-32	Sequence 32, Appli
29	100.5	4.9	1098	4	US-09-268-347-44	Sequence 44, Appli
30	100.5	4.9	1098	4	US-09-797-862-32	Sequence 32, Appli
31	100.5	4.9	2353	3	US-09-377-155-33	Sequence 33, Appli
32	100.5	4.9	2353	3	US-08-913-942-4	Sequence 4, Appli
33	100.5	4.9	2353	4	US-09-669-974-33	Sequence 33, Appli
34	100.5	4.9	2353	4	US-09-797-862-33	Sequence 33, Appli
35	100.5	4.9	2411	4	US-09-268-347-36	Sequence 36, Appli
36	99	4.8	2354	4	US-09-268-347-47	Sequence 47, Appli
37	93	4.7	431	4	US-09-543-681A-5397	Sequence 5397, Ap
38	93	4.5	1117	2	US-08-843-530B-33	Sequence 33, Appli
39	92	4.5	447	4	US-09-543-681A-6231	Sequence 6231, Ap
40	90.5	4.4	1226	1	US-08-280-443-2	Sequence 2, Appli
41	90.5	4.4	1226	1	US-08-457-459-2	Sequence 2, Appli
42	90.5	4.4	1226	1	US-08-555-678-2	Sequence 2, Appli
43	90.5	4.4	1226	5	PCT-US95-02275-2	Sequence 2, Appli
44	90	4.4	512	4	US-09-489-039A-12836	Sequence 12836, A
45	90	4.4	944	4	US-09-134-001C-4352	Sequence 4352, Ap

ALIGNMENTS

RESULT 1
US-09-336-447A-11
; Sequence 11, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: MNCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-11

Alignment Scores:
Pred. No.: 3.6e-13 Length: 610
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-11 (1-610)

Qy 476 AACGATTTAAGACACGATTTTAAATTAAGATCTTGTGACGACGATTTTCAAAATAAAA 535

Db 471 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 490

```
Qy 536 CAAATATTGATCTACTATAAGTAAATATTACTAGAACTGGGTACTTATTATGATGAT--- 592
Db 491 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 510
Qy 593 TCTTATCGTATGATGGAACAAATACAT-----AATATCAATAAG 634
Db 511 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 530
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 531 LeuAspSerLysValGluAsnGlyMetAlaAlaLeuSerGlyLeuPheGln 550
Qy 695 CCAATGCTAGGCAAAACGAGCGTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 551 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrLysGlySerLysSer 570
Qy 755 GCATTAGCCATTGGTGTGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 571 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 590
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 591 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 610

RESULT 2
US-09-336-447A-7
; Sequence 7, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-7

Alignment Scores:
Pred. No.: 3.63e-13 Length: 624
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-7 (1-624)
Qy 476 AACGATTTAAGACACGAGTTTAAATTAAGTTCTTGATGACGATTTCCAAAATAAA 535
Db 485 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 504
Qy 536 CAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACTTATTATGATGAT--- 592
Db 505 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 524
Qy 593 TCTTATCGTATGATGGAACAAATACAT-----AATATCAATAAG 634
Db 525 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 544
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 545 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 564
```

```
Qy 695 CCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 565 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrLysGlySerLysSer 584
Qy 755 GCATTAGCCATTGGTGTGCTCGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 585 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 604
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 605 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 624

RESULT 3
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Alignment Scores:
Pred. No.: 4.15e-13 Length: 889
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-15 (1-889)
Qy 476 AACGATTTAAGACACGAGTTTAAATTAAGTTCTTGATGACGATTTCCAAAATAAA 535
Db 750 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 769
Qy 536 CAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACTTATTATGATGAT--- 592
Db 770 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 789
Qy 593 TCTTATCGTATGATGGAACAAATACAT-----AATATCAATAAG 634
Db 790 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 809
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 810 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 829
Qy 695 CCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 830 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrLysGlySerLysSer 849
Qy 755 GCATTAGCCATTGGTGTGCTCGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 850 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 869
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 870 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 889
```


Db 1942 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1961
Qy 629 AATAAGTCTGCTAAGAAATGCAAACTGGTTTACCCAAACCAATCAGCATTGCTCTATGTTA 688
Db 1962 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1981
Qy 689 GTGCAACCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGAT 748
Db 1982 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 2001
Qy 749 AAAACTGATAGCATTGGTGTGGCTCAGCATTACTGATCGCTTTACCGCTAAAGCG 808
Db 2002 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2018
Qy 809 GGTGTA-----CGCTCAATACCTACAAATCGCGGCATGCTTATGCTGCTCTGTT 859
Db 2019 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2038
Qy 860 GGTATGCAATTC 871
Db 2039 GlyPheSerPhe 2042

RESULT 9

US-09-077-098A-7
; Sequence 7, Application US/09077098A
; Patent No. 6544519

GENERAL INFORMATION:

APPLICANT: TOKUNAGA, Bijl
SAKAGUCHI, Masashi
MATSUO, Kazuo
HAMADA, Fukusaburo
TOKIYOSHI, Sachio

TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS

PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/077,098A

FILING DATE: 19-May-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP97/03222

FILING DATE: 12-SEP-1997

APPLICATION NUMBER: JP 27,148/1996

FILING DATE: 19-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: KORNEAU, Anne M.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TOKUNAGA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2039 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-077-098A-7

Alignment Scores:

Pred. No.: 0.000525 Length: 2039
Score: 114.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.57% Indels: 6
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-077-098A-7 (1-2039)

Qy 569 GAACCTGGTACTTATTAGATGATCTTATCGTATGATGGAACAAATACACATAATATC 628
Db 1939 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1958
Qy 629 AATAAGTCTGCTAAGAAATGCAAACTGGTTTACCCAAACCAATCAGCATTGCTCTATGTTA 688
Db 1959 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1978
Qy 689 GTGCAACCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGAT 748
Db 1979 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 1998
Qy 749 AAAACTGATAGCATTGGTGTGGCTCAGCATTACTGATCGCTTTACCGCTAAAGCG 808
Db 1999 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2015
Qy 809 GGTGTA-----CGCTCAATACCTACAAATCGCGGCATGCTTATGCTGCTCTGTT 859
Db 2016 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2035
Qy 860 GGTATGCAATTC 871
Db 2036 GlyPheThrPhe 2039

RESULT 10

US-09-336-447A-13
; Sequence 13, Application US/09336447A
; Patent No. 6310190

GENERAL INFORMATION:

APPLICANT: HANSEN, ERIC J.

APPLICANT: AEBI, CHRISTOPH

APPLICANT: COPE, LESLIE D.

APPLICANT: MACIVER, ISOBEL

APPLICANT: FISKE, MICHAEL J.

APPLICANT: FREDENBURG, ROSS A.

TITLE OF INVENTION: USPAI AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS

FILE REFERENCE: AMCY:024

CURRENT APPLICATION NUMBER: US/09/336,447A

CURRENT FILING DATE: 1999-06-21

NUMBER OF SEQ ID NOS: 98

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 13

LENGTH: 873

TYPE: PRT

ORGANISM: Moraxella catarrhalis

US-09-336-447A-13

Alignment Scores:

Pred. No.: 0.00109 Length: 873
Score: 110.00 Matches: 42
Percent Similarity: 47.37% Conservative: 30
Best Local Similarity: 27.63% Mismatches: 52
Query Match: 5.37% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-13 (1-873)

Qy 428 CTTGGATTAATCGGCAGCAGCTGAATTCGAATTTGTATAGTTATTTTACGATTTAAGA 487
Db 745 IleAlaThrAsnLysGlnGluLeu-----IleLeuGln 755
Qy 488 CAGGATTTAAATTAAGGTTCTTGATGCAGCTATTTCACAAATAAACAAATATTGAT 547
Db 756 HisAsp---ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 774

```
QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATTAGATGATTTCTTATCGTATGATG 607
Db 775 -----GlnLeuGly-----TyrAlaLeuLys 781
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db 782 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 801
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGC 718
Db 802 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 821
QY 719 GTTCTCTGCTGCGGTAGGATTTAGAGATAAACTGCAATAGCCATTGGT---GTGCGC 775
Db 822 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 841
QY 776 TCACCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATCCTACATGCGC 835
Db 842 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 860
QY 836 GGCATGTCCTATGGTCTCTGCTGTTGGTTATGAAATTC 871
Db 861 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 872

RESULT 11
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Alignment Scores:
Pred. No.: 0.00185 Length: 892
Score: 108.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.27% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-5 (1-892)
QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTTGATAGTATTATTAAACGATTAAAGA 487
Db 764 IleAlaThrAsnLysGlnLeuIleLeuGln-----AsnAsp----- 776
QY 488 CAGCATTTAAATTAAGTTCTTGATGCACGTATTTCCAAAATAAACAATAATTCAT 547
Db 777 -----ArgLeuAsnGlnIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 793
QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATTAGATGATTTCTTATCGTATG 607
Db 794 -----GlnLeuGly-----TyrAlaLeuLys 800
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db 801 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 820
```

```
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGC 718
Db 821 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 840
QY 719 GTTCTCTGCTGCGGTAGGATTTAGAGATAAACTGCAATAGCCATTGGT---GTGCGC 775
Db 841 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 860
QY 776 TCACGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATCCTACATGCGC 835
Db 861 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 879
QY 836 GGCATGTCCTATGGTCTCTGCTGTTGGTTATGAAATTC 871
Db 880 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 891

RESULT 12
US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Alignment Scores:
Pred. No.: 0.00234 Length: 831
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-1 (1-831)
QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTTGATAGTATTATTAAACGATTAAAGA 487
Db 703 IleAlaThrAsnLysGlnGluLeuIleLeuGln-----AsnAsp----- 715
QY 488 CAGCATTTAAATTAAGTTCTTGATGCACGTATTTCCAAAATAAACAATAATTCAT 547
Db 716 -----ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 732
QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATTAGATGATTTCTTATCGTATG 607
Db 733 -----GlnLeuGly-----TyrAlaLeuLys 739
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db 740 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 759
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGC 718
Db 760 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 779
QY 719 GTTCTCTGCTGCGGTAGGATTTAGAGATAAACTGCAATAGCCATTGGT---GTGCGC 775
Db 780 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 799
```


Qy 776 TCACGCATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTCAATAGGC 835
Db 800 LeuSerAspThrGlyLysSerThrTyrllysleGlyLeuSerTrpSerAsp---AlaGly 818
Qy 836 GGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 819 GlyLeuSerGlyGlyValGlyGlySerTyrlArgTrp 830

RESULT 13

US-09-540-236-3401
; Sequence 3401, Application US/09540236
; Patent No. 6673910

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3401
; LENGTH: 878
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3401

Alignment Scores:
Pred. No.: 0.00239 Length: 878
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-540-236-3401 (1-878)

Qy 428 CTTGGAATAATCGGAGCAGCTGAAATTTGATATGATATTTTAAAGATTAAAGA 487
Db 750 IleAlaThrAsnLysGlnGluLeuLeuGln-----AsnAsp----- 762
Qy 488 CACGATTTAAATATAAAGTCTTGATGCAGCTATTTCCAAATAATAACAAATATTGAT 547
Db 763 -----ArgLeuAspArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 779
Qy 548 ACTATAAGTAATATTACTAGAACTCGGACTTTATTAGATGATTTCTTCGTATGATG 607
Db 780 -----GlnLeuGly-----TyrAlaLeuLys 786
Qy 608 GAACAAATACATAATATCAATAAGTTGTCTTAA-----GAATGCAAACT-----GGT 658
Db 787 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 806
Qy 659 TTAGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGCCAAACCGAGC 718
Db 807 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 826
Qy 719 GTTTCCTCGGTAGGAGTGTATAGAGATAAACTGCATTAGCCATTGGT---GTCCGC 775
Db 827 ValLeuPheGlySerGlyTyrlHisAsnGlyGlnAlaAlaValSerLeuGlyAlaAlaGly 846
Qy 776 TCACGCATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTCAATAGGC 835
Db 847 LeuSerAspThrGlyLysSerThrTyrllysleGlyLeuSerTrpSerAsp---AlaGly 865
Qy 836 GGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 866 GlyLeuSerGlyGlyValGlyGlySerTyrlArgTrp 877

RESULT 14

US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190

GENERAL INFORMATION:

; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

Alignment Scores:

Pred. No.: 0.00245 Length: 941
Score: 107.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.22% Indels: 18
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-9 (1-941)

Qy 437 AATCGGCGAGCAGCTGAAATTTGATATGATATTTTAAACGATTTTAAAGACACGATTTT 496
Db 788 AsnLysThrAlaIleGluGlnAsnIleAsn-----ArgThrValAlaAsnGlyPhe 804
Qy 497 AAATAAAGTCTTGATGCAGCTATTTCCAAATAATAACAAATATT----- 544
Db 805 GluIleGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuIleLeuGlnAsn 824
Qy 545 GATACATAAGTAATAATTTACTAGAACTGGGTACTTATTATGATGAT----- 592
Db 825 AspArgLeuAsnGln---IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 843
Qy 593 ---TCTTATGATGGAACAAATAATACATAATATCAATAGTTGTCTTAAA---GAA 646
Db 844 LeuGlyTyrlAlaLeuLysGluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 863
Qy 647 TTCCAAACCT-----GTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAAT 700
Db 864 ArgGlnThrAlaGlyGlyIleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSer 883
Qy 701 GGTGTAGGCAAAACGAGCGTTTCTGTCGCGTAGGAGGTATAGAGATAAACTGCATTA 760
Db 884 ArgAlaGlyGluHisValLeuPheGlySerGlyTyrlHisAsnGlyGlnAlaAlaVal 903
Qy 761 GCCATTGGT---GTCCGCTCACCGCATTTACTGATCGCTTTTACCGCTAAAGCGGTGTAGCG 817
Db 904 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrlLysIleGlyLeuSer 923
Qy 818 TTCAAATACCTACCAATGCGGCATGCTTATGTCGTCCTCTGTTGGTTATGAATTC 871
Db 924 TrpSerAsp---AlaGlyGlyLeuSerGlyGlyValGlyGlySerTyrlArgTrp 940

RESULT 15

US-09-268-347-24
; Sequence 24, Application US/09268347
; Patent No. 6335182

GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268.347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 24
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-24

Alignment Scores:
Pred. No.: 0.00549 Length: 1002
Score: 104.00 Matches: 30
Percent Similarity: 54.22% Conservative: 15
Best Local Similarity: 36.14% Mismatches: 36
Query Match: 5.08% Indels: 2
DB: 4 Gaps: 2

US-10-030-529A-1 (1-1168) x US-09-268-347-24 (1-1002)

QY	626	ATCAATAGTTGTCTAAAGATTGCCAACTGGTTAGCCCAACCAATCAGCATTGTTCTATG	685
DB	921	ValAsnLysValGlyLysArgAlaAspAlaGlyThrAlaSerAlaLeuAlaAlaSerGln	940
QY	686	TTAGTGCACCAACCAATGGTAGGCAAAACGAGCGTTTCTGTCGGTAGGAGGTTATAGA	745
DB	941	LeuProGlnAlaSerMetSerGlyLysSerMetValSerIleAlaGlySerSerTyrGln	960
QY	746	GATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTACTATCGCTTTACCGCTAAA	805
DB	961	GlyGlnSerGlyLeuAlaIleGlyVal---SerArgIleSerAspAsnGlyLysValIle	979
QY	806	CGGGGTGTAGGTTCAATACCTACAATGGCGGCATGCT---TATGGTCTTCTGTTGGT	862
DB	990	IleArgLeuSerGlyThrThrAsnSerGlnGlyLysThrGlyValAlaAlaGlyValGly	999
QY	863	TATGAATTC	871
DB	1000	TyrGlnTrp	1002

Search completed: May 13, 2004, 08:30:41
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 15, 2004, 02:41:15 ; Search time 3518 Seconds
(without alignments)
9914.448 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 1168
Sequence: 1 ataataacgtcattgacatt.....aagccgttaaggtggac 1168

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_estovl:*
6: em_estovl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estfun:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	66.8	5.7	1101 29	CNS0039G
2	62	5.3	1200 13	BX437758
3	61.6	5.3	442 29	CNS00H2T
4	59.8	5.1	1151 28	AF416115

C 5	58.8	5.0	747	29	CNS011RQ
C 6	58.8	5.0	1101	29	CNS0039G
C 7	57.2	4.9	1182	13	BX338616
8	56.8	4.9	576	29	CNS035N7
9	56.8	4.9	829	29	CNS011NU
C 10	56.8	4.9	1101	29	CNS00GDR
C 11	56.6	4.8	1165	13	BX338369
C 12	56	4.8	759	29	CNS06QXV
C 13	55.4	4.7	1044	13	BX415231
C 14	55.2	4.7	1101	29	CNS012UN
C 15	55.2	4.7	1101	29	CNS017YG
C 16	55	4.7	950	13	BX415411
C 17	55	4.7	1101	29	CNS012I9
C 18	55	4.7	1201	9	AL524807
C 19	55	4.7	1201	9	AL536104
C 20	54.6	4.7	351	12	BM168097
C 21	54.6	4.7	482	12	BM168129
C 22	54.6	4.7	783	29	CNS00A1S
C 23	54.6	4.7	890	28	AQ026918
C 24	54.6	4.7	1200	13	BX436510
C 25	54.4	4.7	865	28	BH180441
C 26	54.4	4.7	865	29	CNS07MAN
C 27	54	4.6	935	28	B10881
C 28	54	4.6	982	13	BX349562
C 29	54	4.6	989	29	CNS015PZ
C 30	54	4.6	1201	9	AL532464
C 31	53.8	4.6	1073	10	BF274651
C 32	53.8	4.6	1200	13	BX437739
C 33	53.6	4.6	942	29	CNS018GS
C 34	53.4	4.6	1101	29	CNS00EVL
C 35	53.2	4.6	523	29	CE542343
C 36	53.2	4.6	961	14	CK230311
C 37	53	4.5	1101	29	CNS00210
C 38	53	4.5	1201	13	BX402521
C 39	52.8	4.5	834	29	CNS00868
C 40	52.8	4.5	1180	13	BX436369
C 41	52.8	4.5	1225	29	CNS0161D
C 42	52.6	4.5	816	29	CNS02KJV
C 43	52.4	4.5	908	29	AG160668
C 44	52.4	4.5	1101	29	CNS017V2
C 45	52.4	4.5	1200	29	CNS016CO

ALIGNMENTS

RESULT 1
CNS0039G
LOCUS
DEFINITION

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL063921 GI:4941778

GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of


```

506 GTTCTGATGACGATATTCACAAATAAACAATAATGATGACTATAAGTAATAATTTA 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
690 RKRRATWTTWAANAWAGAAARGAGAGRAATTTTTTTATKAGARRGATWTTAW 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 CTAGAACTGGGACTATTTAGATGATCTTCATCGTATGATGGAACAAAATACACATAAT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 WTATPARGAGADTTTTTTATPAMWTATTTTTTWWAAGATDKAAAAAAMWTTTWTAAAA 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 ATCAATAA 633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
810 AAAATTWW 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
LOCUS      CNS00H2T
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35M12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL074554
VERSION     AL074554.1 GI:4953930
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 442)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osogawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES             Location/Qualifiers
     source           1..442
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR35M12"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN
Query Match      5.3%; Score 61.6; DB 29; Length 442;
Best Local Similarity 46.3%; Pred. No. 0.045;
Matches 130; Conservative 18; Mismatches 133; Indels 0; Gaps 0;

Qy 862 TTATGAATTCATCATGCTTTAATCAATCGCTTATTAAGCTTATCAATCAGTCTTCTACT 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 TTWTTAATTTATTTTTTTTTTTTATATATATTTTTTTGATTTTTTTTATATWTTTW 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 922 ATGTTCTCTCATATTGACCTTTCTATTATCTTTGTTATAGCTTTTGCTGTATAA 981
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 WTTTTCNTAAATTTTATTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 982 AACCGTTTTTATGACCTTTTATTAATAAAGCTTTTAAAGCTTATCAATCAGTCTTCTACT 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 ATWATTTTTTAAWATATTTATTTATTTATTTTTTTTTTTTAAATTTTTTGTTTTWTW 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Qy 1042 TCACTTTTTTCACATATTATCCGCCACTTCTAAACCGTAATATTAGTTGGTTAGCC 1101
   : : ||| ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||
Db 326 TTWTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1102 TAAATTCGGTAGCTTCATCGGAATTTTTTTCTAAATGTTCT 1142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 TTWTTWTTTTTTTTTTTTTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
LOCUS      AF416115/c
DEFINITION AF416115
            Haemophilus influenzae biotype aegyptius genomic, genomic survey sequence.
ACCESSION  AF416115
VERSION     AF416115.1 GI:18034818
KEYWORDS   GSS.
SOURCE     Haemophilus influenzae biotype aegyptius
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
ORGANISM   Haemophilus influenzae biotype aegyptius
            Haemophilus influenzae biotype aegyptius
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE   1 (bases 1 to 1151)
AUTHORS    Smoot, L.M., Franke, D.D., McGillivray, G. and Actis, L.A.
TITLE      Genomic analysis of the F3031 Brazilian purpuric fever clone of
            Haemophilus influenzae biogroup aegyptius by PCR-based subtractive
            hybridization
JOURNAL    Infect. Immun. 70 (5), 2694-2699 (2002)
MEDLINE    21950594
PUBMED     11953414
COMMENT    Contact: Actis LA
            Microbiology
            Miami University
            40 Pearson Hall, Oxford, OH 45056, USA
            Email: actisla@muhio.edu
            associated with Brazilian purpuric fever biotype: aegyptius
            isolate = F2
            Class: PCR-based subtractive hybridization.

FEATURES             Location/Qualifiers
     source           1..1151
                     /organism="Haemophilus influenzae biotype aegyptius"
                     /mol_type="genomic DNA"
                     /strain="F3031"
                     /db_xref="taxon:725"
                     /clone_lib="Haemophilus influenzae biotype aegyptius
                     F3031"

ORIGIN
Query Match      5.1%; Score 59.8; DB 28; Length 1151;
Best Local Similarity 63.6%; Pred. No. 0.076;
Matches 91; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 627 TCAATAAGTTGTCTAAAGAATTCGAAACTGGTTTAGCCAACTACATGCTATGT 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 TAATAACCTTANTAAGATCTTAACGTGCTCTTCTGCTCAAGCTGCATTAATGTT 86
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 687 TAGTCAACCAATGGTGTAGGCAACGACGCTTCTGCTCGGTAGGAGGTTATAGAG 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 TATTCCAACCGTAAACGTAGGTAAATTAATCTTACTGCTGTAGGTGGTTATAAAT 26
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 747 ATAAACTGCATTAGCCATTGGT 769
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 CTCAAATGCAGTTGCTGTAGGT 3
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LOCUS      CNS011RQ/c
DEFINITION CNS011RQ
            Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN06A03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL100640
VERSION     AL100640.1 GI:5612251

```


[illegible]

RESULT 8	CNS035N7	576 bp	DNA	linear	GSS 01-SEP-2000
LOCUS	Tetradon nigroviridis genome survey sequence PUC-Ori end of clone				
DEFINITION	214A06 of library G from Tetradon nigroviridis, genomic survey sequence.				
ACCESSION	AL228940				
VERSION	AL228940.1	GI:7887933			
KEYWORDS	GSS: genome survey sequence.				

ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acantopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE	2
AUTHORS	Rozest Crollius H., Jaillon O., Dasilva C., Ozouf-Costaz, C., Fzimes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot A. and Weissenbach, J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Genome Res. 10 (7), 939-949 (2000)
MEDLINE	20359837
PUBMED	10899143
REFERENCE	3 (bases 1 to 576)

3 (pages 1 to 376)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91060 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

- Web : www.genoscope.cns.fr/
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

```

FEATURES
    source
        Location/Qualifiers
            1..576
                /organism="Tetraodon nigroviridis"
                /mol_type="genomic DNA"
                /db_xref="taxon:99883"

```

```

/clone="214A06"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG214BA03SP1-end :
prc-ori"

```

ORIGIN	Query Match	Score	DB	Length
FUC-ORI	4.9%	56.8	29	576

Query Match 4.9%; Score 56.8; DB 29; Length 576;


```

Db      987  THMAKHTTTTMMKTTAGYCDKVMHMHHTKMMGTMMVCMHMYNTHHHHWC 928
Qy      289  GCCAAAGAAGTGGATTTCTAAACAGGCTACTTATCTTGAATTA CAGCAATATATGCGCTTA 348
Db      927  ATMCATKTMKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 868
Qy      349  TACTCTCTGTTCTCGTGACATATGCTCCCTGGGTTTCTCCCTAGCCCTACTGTTATATCC 408
Db      867  SMKYATMKMHHTMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMM 808
Qy      409  GATGCTGATCCTGATCAACTTGAATAAATCGGACGAGCTGAAATTTGATATAG 468
Db      807  KKSIVAGKAAAMTATTAHAWATTTTAYATHWTKAKYTKYKTWAMWMMGMKAKTMM 748
Qy      469  TTATTTTAAAGATTAGACAGATTTTAAATTTAAAGTTTCTTGATGACGATATTTCCAA 528
Db      747  AKAYAACWMAATTTTACATWYAGATHAAAWMAWAATTTCTWKAKAHAAKHAATTTGT 688
Qy      529  AAATAAACAATAATGATAGTAAATATTTTACTAGAACCTGGGTACTTATTTAGA 588
Db      687  TTCCACACATAATWNTTTTTTAWACAAHTTTTTTBAATAAWAATKAKWYATAATWYAYA 628
Qy      589  TGATTTCTTATCGTATGATGGAACAAATACACATAATCAATAAGTTGTCTAAAGAAAT 648
Db      627  THACTARATTTAAKWAATAATTTTTTTTAAKAMATATACHTTTTABWTAFTAKTKATAMAK 568
Qy      649  GCAA 652
Db      567  TCTA 564

RESULT 11
BX338369/c
LOCUS
DEFINITION BX338369 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI058YF14 5-PRIME, mRNA sequence.
ACCESSION BX338369
VERSION BX338369.1 GI:30343682
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1165)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI058DC07QPL.
FEATURES
Location/Qualifiers
1. 1165
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 4.8%; Score 56.6; DB 13; Length 1165;
Best Local Similarity 37.4%; Pred. No. 0.32; Indels 1; Gaps 1;
Matches 218; Conservative 57; Mismatches 307;

```

```

Qy      469  TTATTTTAAAGATTAGACAGATTTTAAATTTAAAGTTTCTTGATGACGATTTCCAA 528
Db      1165  WWWTTTTTAAAWMTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1106
Qy      529  AAATAAACAATAATGATAGTAAATATTTTACTAGAACCTGGGTACTTATTTAGA 588
Db      1105  TTWTATWATAWMTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATATK 1046
Qy      589  TGATTTCTTATCGTATGGAACAAATACACATAATCAATAAGTTGTCTAAAGAAAT 648
Db      1045  TTTTTTTTATWATTTTAAATAAAATWAAAAAATAATWAAAAAATAATATATATAWAA 986
Qy      649  GCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTAGTGTAGTGCACCAATGCTGTAGG 708
Db      985  ATATATATWMTTWWAAAAATWAAATWAAATWAAATWAAATWAAATWAAATWAAATWAA 926
Qy      709  CAAAAACGAGCGTTTCTGCTCGGTAGGAGGTATPAGAGATAAAACCTGCATTTAGCCAT 768
Db      925  TAAAAW-ATWTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 867
Qy      769  TGTGCGGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGGTAGCGTTCAATACCTA 828
Db      866  AAWAAWATTTTWTWATAWATWATWATWATWATWATWATWATWATWATWATWATWATWAT 807
Qy      829  CAATGGCGGATCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
Db      806  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 747
Qy      889  CACTAATCGTTTGGTTTATATAAAAGGCTAAATGTTTCTCTCCATCAGATTTAGCCCTTCT 948
Db      746  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 687
Qy      949  TATTATCTTTGTTATAGCTTTTGTGCTGCTTATAAACCCTTTTTTTAGCCACTTTTATTAAT 1008
Db      686  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 627
Qy      1009  TAAGCTTTTAAAGCTATTCAATCAGTTCTACTTCTACTTCTACTTCTACTTCTACTTCT 1051
Db      626  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 584

RESULT 12
CNS06QXV/c
LOCUS
DEFINITION T7 end of clone AWOAA009H09 of library AWOAA from strain CLIB 89 of
Yarrowia lipolytica, genomic survey sequence.
ACCESSION AL411257
VERSION AL411257.1 GI:12180512
KEYWORDS GSS.
SOURCE Yarrowia lipolytica
ORGANISM Yarrowia lipolytica
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodascaceae; Yarrowia.
1 (bases 1 to 759)
Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Boloitin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Iloriente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekaiia,P., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED 11152876
2 (bases 1 to 759)
Casaregola,S., Neuvéglise,C., Lepingle,A., Bon,E., Feynrol,C.,
Artiguenave,F., Wincker,P. and Gaillardin,C.
Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
lipolytica
FEBS Lett. 487 (1), 95-100 (2000)
20584727

```


collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> -. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

```

FEATURES
    pbe02bac11: Location/Qualifiers
        1..1101
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACN08024"
            /clone_lib="DrosBAC"
            /plasmid="pbe02BAC11"
            /note="end : T7"

```

```

ORIGIN                                     Query Match      4.7%; Score 55.2; DB 29; Length 1101;
Best Local Similarity 43.1%; Pred. No. 0.61;
Matches 132; Conservative 23; Mismatches 151; Indels 0; Gaps 0;
Notes=End : 17.

844 TTATGGTGCTTCGTGTTATGAATCTAATCATTAACGTTTAACTACTAATCGTTTTGG 903
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
558 TTTCTWYTWWTWTTTCCCTTTTCTTTTTTTTTTTTTWWTTTTATTATTATVTCCTT 499
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
904 TTATAATAAAAAGGCATAATGTTTCTCCTCACATTTAGCCCTTCTTATTATCTTGTGA 963
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
498 TTTATACTTWWTTTYATTTTCTTYYTTTTTTTTTTTTTTTTTTTTTTTTTTCCTT 439
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
964 TAGCTTTTGTGCTTTATAAAAACCGTTTTTTTAGCCACTTTATTAATTAAGCTTTTAAGCCT 1023
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
438 TTGTTTTTYYKATTTTTTTTWWTTTWTTTTTTTTTTTTTTTTTTTTTTTTATTTTTT 379
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1024 ATTCAATCAAGTCTCACTTTTCACTTTTTCACCATAATTATCCGCCACTTTCAAACGGTAA 1083
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
378 TTTTTTTTTTTTWTTTTTTTTTTTTTTTTTTTTTTTCWTKWTTTWTGTTYTTWTTTTTTTCWT 319
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1084 TATTAAAGTGGTTTAGCCTAAATTTGGGTACCTTCTATCGGAATTTTTCTAAATGTTCTA 1143
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
318 TKTWTTTTTTTTTAWTTTTTAWTTTTTWTATACCTTTTCTTCTTTTTTATACTYTITW 259
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1144 AAATTA 1149
   | | | | |
258 ATWTTW 253
   | | | | |

```

RESULT 15	CNS017YG	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS017YG				
DEFINITION	Drosophila melanogaster genome survey sequence S86 end of BAC				
	BAC37P24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL108658				
VERSION	AL108658.1	GI:5628962			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
	Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr				
	- Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton				

and Genevieve Payan. It has been constructed in the vector pBelobAC11.

```

FEATURES
source
pbac000001
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37P24"
/clone_lib="DrosBAC"
/plasmid="pBel0BAC11"
/note="end : Sp6"

```

ORIGIN

Query Match	4.7%;	Score 55.2;	DB 29;	Length 1101;
Best Local Similarity	38.7%;	Pred. No. 0.61;		
Matches 111;	Conservative 38;	Mismatches 138;	Indels 0;	Gaps 0;
Qy	843	CTTATGGTCTCTGTTGGTTATGAATTCCTAACTCAATACGTTTAAACACACTAAATCGTTTTCG	902	
Db	189	CAWATWTTTCTCTATNTATATATTTTTNNCTATWTTWTTNNTTTTTTTTTWWMTTWT	248	
Qy	903	GTATATAATAAAGGCTAAATGTTTCTCCTCACATTTAGCCTTCTTATTTATCTCTTGT	962	
Db	249	WTTTWTWTTTATTTTWTICANTTTTTWWYCTTHAAATCWAATTTTTCATTTTTTTTTT	308	
Qy	963	ATAGCTTTTGTCTTATAAAACCGTTTTTTTAGCCACTTTTATTAATTAAGCTTTTAAAGCC	1022	
Db	309	WCTACITTTATATTAATAWNNWCCTCATWTTTCTAAATTTTNCATWTTTWTWCCATTY	368	
Qy	1023	TATTCATCAGTTTCTACTTTTTCACCATATTTATCCGCCACTTCTAAAAACGGTA	1082	
Db	369	TYATWTTTNTTNNANATWTTWTTATTTTWAYATTTTTTCCATCAACCAACATTTACH	428	
Qy	1083	ATATTAAGTTGGTTTAGCCATAATGGGTACCTTCTATCGGAATTT	1129	
Db	429	ATACTAWYCTTTTATWCAACVYTTTACATAWNNWCCTCTATHAACCTCT	475	

Search completed: May 15, 2004, 05:20:56
Job time : 3528 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 04:22:11 ; Search time 597 Seconds
(without alignments)
8878.461 Million cell updates/sec

Title: US-10-030-529a-1

Perfect score: 1168

Sequence: 1 ataatacgtcattgacatt.....aagccgttaagtgaggac 1168

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.2	4.9	8139	13	US-10-282-1222a-30886
2	54	4.6	1830121	15	Sequence 30886, A
3	54	4.6	1830121	16	Sequence 1, Appli
4	53.2	4.6	2596	10	Sequence 4, Appli
5	53.2	4.6	2673	10	Sequence 12, Appl
6	53.2	4.6	4228	10	Sequence 16, Appl
7	53.2	4.6	94750	13	Sequence 38, Appl
8	52.4	4.5	3295	10	Sequence 8, Appli
9	51.8	4.4	18624	15	Sequence 1675, Ap
10	50.8	4.3	9117	15	Sequence 1041, Ap
11	50.4	4.3	3673778	15	Sequence 1, Appli
12	50	4.3	673	13	Sequence 57895, A
13	49.2	4.2	3673778	15	Sequence 2, Appli
14	49	4.2	18683	15	Sequence 286, App

15	49	4.2	18683	15	US-10-240-452-34	Sequence 34, Appl
16	48.6	4.2	5979	15	US-10-239-676-17	Sequence 17, Appl
17	48.6	4.2	5979	15	US-10-240-453-25	Sequence 25, Appl
18	48	4.1	9964	15	US-10-311-455-71	Sequence 71, Appl
C 19	48	4.1	3673778	15	US-10-312-841-2	Sequence 2, Appli
20	47.6	4.1	14950	15	US-10-311-455-1230	Sequence 2319, Ap
21	47.2	4.0	413	9	US-09-960-352-2919	Sequence 1230, Ap
22	47.2	4.0	431	9	US-09-960-352-5558	Sequence 5558, Ap
23	47.2	4.0	9539	15	US-10-239-676-51	Sequence 51, Appl
24	47.2	4.0	9539	15	US-10-240-453-53	Sequence 53, Appl
C 25	47	4.0	1872	13	US-10-424-599-140661	Sequence 140661,
26	47	4.0	6203	15	US-10-239-676-184	Sequence 184, App
27	47	4.0	6203	15	US-10-240-453-274	Sequence 274, App
28	47	4.0	6767	13	US-10-221-714A-330	Sequence 330, App
29	46.8	4.0	428	9	US-09-960-352-573	Sequence 573, App
30	46.6	4.0	446	9	US-09-960-352-3400	Sequence 3400, Ap
31	46.2	4.0	5989	13	US-10-221-613-172	Sequence 172, App
32	46.2	4.0	5989	15	US-10-240-452-52	Sequence 52, Appl
33	46.2	4.0	6171	15	US-10-311-455-761	Sequence 761, App
34	46.2	4.0	6923	13	US-10-221-613-157	Sequence 157, App
35	46.2	4.0	271990	15	US-10-195-144-87	Sequence 87, Appl
36	46.2	4.0	271990	16	US-10-345-072-87	Sequence 87, Appl
37	46	3.9	430	9	US-09-960-352-5818	Sequence 5818, Ap
38	46	3.9	5992	13	US-10-221-613-168	Sequence 168, App
39	46	3.9	6664	13	US-10-221-613-331	Sequence 331, App
40	46	3.9	9483	13	US-10-221-613-47	Sequence 47, Appl
41	46	3.9	9483	15	US-10-311-455-349	Sequence 349, App
42	45.8	3.9	5501	13	US-10-221-613-203	Sequence 203, App
43	45.8	3.9	5501	15	US-10-204-708-37	Sequence 37, Appl
44	45.8	3.9	5501	17	US-10-240-589C-63	Sequence 63, Appl
45	45.8	3.9	9110	15	US-10-311-455-2402	Sequence 2402, Ap

ALIGNMENTS

RESULT 1

US-10-282-1222a-30886
; Sequence 30886, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30886
; LENGTH: 8139
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-10-282-122A-30886

Query Match          4.9%; Score 57.2; DB 13; Length 8139;
Best Local Similarity 51.6%; Pred. No. 0.15;
Matches 131; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 621 ATATATCAATAAGTGTCTTAAGAAATGCAAACTGGTTTAGCCCAACCAATCAGCATGTG 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7886 ATCATATCAATAAAGTGGATAAAGATCTCCGTCGGGTATCGCGGTGCCACAGCGGTG 7945

Qy 681 CTATGTTAGTCGAACCAATGGTGTAGGCAAAACGAGGTTTCTGCTCGGTAGAGGTT 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7946 CGTCTTACACGGTCCTCAATGAAGCAGGTAAGAGCATGTGTTCTCTTGTGTAGGTCTT 8005

Qy 741 ATAGAGATAAAACTGATTAGCCATTGGTGTGGCTCACGCATTACTGCTTTTACCG 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8006 ATCGTAGTGATCTGCTATTGCTTTCAGTTGGTTATGCAAGAACTCAGATAACATAAGATCT 8065

Qy 801 CTAAGCGGGTGTAGCGTTCAATACCTACAATGGCGCATGCTTTAGTGGTCTCTGTG 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8066 CAATCAAACTTGGTGTGGTGTATGAACAGCCGCGGTGATGTGAACCTTCGGTGTAGTATCG 8125

Qy 861 GTTATGATTCTAA 874
    ||||| ||||| |||||
Db 8126 GTTATCAATGGTAA 8139

RESULT 2
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PBI86P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40810)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
```

```
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      4.6%; Score 54; DB 15; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 8.7;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCAGTTCTACTTTTCACCTTTTTCACCATATATATCGGCCACTTCTAAACGGTAA 1083
Db 115616 ATTAAAGTTTTCAGCAGCTTTCGCTGTTTATCATATATATCGGCCACTTCAAGATAGTGA 115557
Qy 1084 TATTAAGTTGGTTTAGCCTAAATTTGGTACCTTCTATCGGAATTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTCGCAATCGTGCCTTCATCGGTAATTCCTCTAAGTGTCTA 115497
Qy 1144 AAATTAAGCGGTTAAAGTGCG 1165
Db 115496 AGATTAACCATTAAGTACG 115475

RESULT 3
US-10-329-670-1/c
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; PRIOR FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36836)..(36836)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:


```
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
FEATURE:
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
FEATURE:
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
FEATURE:
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Query Match 4.6%; Score 54; DB 16; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 8.7;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCAGTTCACATTTTTCACATATATATCGGCACATTTCTAAACGGTAA 1083
Db 115616 ATTAAAGTTTGACGACATTCGCTGTTTATCATATATATCGCCACATTCGAATAGTGA 115557

Qy 1084 TATTAAGTTGGTTAGCCTAAATTTGGTACCTTCTATCGGATTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTTGCGAAATCGTGCTTCATCGGGTATTTCTCTAAGTGTCTA 115497

Qy 1144 AAATTAAGCGCTTAAAGTGCG 1165
Db 115496 AGATTAACCATTAAGTAGG 115475

RESULT 4
US-09-952-267-4
; Sequence 4, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2596
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-952-267-4

Query Match 4.6%; Score 53.2; DB 10; Length 2596;
Best Local Similarity 54.0%; Pred. No. 0.6;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGCTAAAGAAATTCGAACTGTTTAGCCAAACCAATCAGCATTTGCTAT 684
Db 2111 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGGCTGCCACGCTGCACTTGGTGGCTA 2170

Qy 625 GTTAGTGCACCAAAATGGTGTAGGCAAAACAGCGTTTCTGCTGCGGTAGGAGTTATAG 744
Db 2180 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGGCTGCCACGCTGCACTTGGTGGCTA 2299

Qy 685 GTTAGTGCACCAAAATGGTGTAGGCAAAACAGCGTTTCTGCTGCGGTAGGAGTTATAG 744
Db 2240 TCTATTCCAGCCTTATAGCGTTGTAAGTTTAAATGCGACGCTGCACTTGGTGGCTA 2299

Qy 745 AGATAAACTGCATTAAGCCATTGGTGTGCGCTCAGCAATTAATGATCGCTTACCGCTAA 804
Db 2300 CTCAAAATCTGCGTTGCTATCGTGTGCTATCGTGTGCTATCGTGTGCTATCGTGTGCTA 2359

Qy 805 AGCGGTGTAGCGTTCAATACC 826
Db 2360 AGCTGTGCGCGGATTAAATACC 2381

RESULT 6
US-09-952-267-16
; Sequence 16, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
```

; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 4228
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-952-267-16

Query Match 4.6%; Score 53.2; DB 10; Length 4228;
Best Local Similarity 54.0%; Pred. No. 0.75;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGCTAAAGAAATTGCAAACTGGTTAGCCAAACCAATCAGCAATGTCTAT 684
| | | | |
Db 3629 TATCACAGCTTTAGACAGATAAAGTTGAAACCGGTATGCTCCCAAGCTGCCCTAAGTGG 3688
| | | | |
Qy 685 GTTAGTGCAACCAAAATGGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATAG 744
| | | | |
Db 3689 TCTATTCAGCTTTATAGCTGTGTAAGTTTAATGCGACCGCTGCACCTTGGTGGCTAIGG 3748
| | | | |
Qy 745 AGATAAAACTCATTAGCAATGGTGTGGCTCAGCAATTAATGATCGCTTTACCGCTAA 804
| | | | |
Db 3749 CTCAAAATCTCGGTTGCTATCGTGTGCTGCTATCGTGTGAATCCAAATCTGGCGTTTAA 3808
| | | | |
Qy 805 AGCGGGTGTAGCTTCAATACC 826
| | | | |
Db 3809 AGCTGGTGGCGGATTAATACC 3830
| | | | |

RESULT 7

US-10-672-787-38
; Sequence 38, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-672-787-38

Query Match 4.6%; Score 53.2; DB 13; Length 94750;
Best Local Similarity 54.0%; Pred. No. 3.2;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGCTAAAGAAATTGCAAACTGGTTAGCCAAACCAATCAGCAATGTCTAT 684
| | | | |
Db 26785 TATCACAGCTTTAGACAGATAAAGTTGAAACCGGTATGCTCCCAAGCTGCCCTAAGTGG 26844
| | | | |
Qy 685 GTTAGTGCAACCAATGGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATAG 744
| | | | |
Db 26845 TCTATTCAGCTTTATAGCTGTGTAAGTTTAATGCGACCGCTGCACCTTGGTGGCTAIGG 26904
| | | | |
Qy 745 AGATAAAACTCATTAGCAATGGTGTGGCTCAGCAATTAATGATCGCTTTACCGCTAA 804
| | | | |
Db 26905 CTCAAAATCTCGGTTGCTATCGTGTGCTATCGTGTGAATCCAAATCTGGCGTTTAA 26964
| | | | |
Qy 805 AGCGGGTGTAGCTTCAATACC 826
| | | | |
Db 26965 AGCTGGTGGCGGATTAATACC 26986
| | | | |

RESULT 8

US-09-952-267-8
; Sequence 8, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3295
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-952-267-8

Query Match 4.5%; Score 52.4; DB 10; Length 3295;
Best Local Similarity 55.5%; Pred. No. 0.98;
Matches 101; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 645 AATTGCAAACTGGTTTAGCCAAACCAATCAGCAATGTCTATGTTAGTCAACCAATGTG 704
| | | | |
Db 2854 AGTTGAAACGGTATGGCTGCCCAAGCTGCCCTAAGTGTCTATTCCAGCTTATAGCG 2913
| | | | |
Qy 705 TAGSCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATPAAAACTGCATTAGCCA 764
| | | | |
Db 2914 TTGTAAGTTTAAATGCGACCGCTGCACCTTGGTGGCTATGGCTCAAAATCTGCGGTGCTA 2973
| | | | |
Qy 765 TTGTTGTGCGTACGCAATTAATGATCGCTTACCGCTAAAGCGGTGTAGCGTTCATA 824
| | | | |
Db 2974 TCGTGTGCTGCTATCGTGTGAATCCAAATCTGGCGTTTAAAGCTGGTGGCGCGATTATA 3033
| | | | |
Qy 825 CC 826
| |
Db 3034 CC 3035
| |

RESULT 9

US-10-311-455-1675
; Sequence 1675, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1675
; LENGTH: 18624
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

Search completed: May 15, 2004, 06:52:15
Job time : 637 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 02:44:00 ; Search time 112 Seconds
(without alignments)
5787.345 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 1168
Sequence: 1 ataatacgttcattgacatt.....aagcgttaaaaggtggac 1168

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.6	4.9	7218	1	US-08-232-463-14
C 2	54	4.6	1830121	4	US-09-557-884-1
C 3	54	4.6	1830121	4	US-09-643-990A-1
4	53.2	4.6	2596	4	US-09-336-447A-4
5	53.2	4.6	2604	4	US-09-540-236-756
6	53.2	4.6	2673	4	US-09-336-447A-12
7	53.2	4.6	4228	4	US-09-336-447A-16
8	53.2	4.6	94750	4	US-09-596-002-38
9	52.4	4.5	3295	4	US-09-336-447A-8
10	45.8	3.9	5501	4	US-10-204-708-37
C 11	45.4	3.9	640681	4	US-09-790-988-1
12	45	3.9	640681	4	US-09-790-988-1
13	44	3.8	11049	4	US-10-204-708-22
14	43.4	3.7	2037	3	US-08-913-942-14
15	43.4	3.7	2079	4	US-09-268-347-25
16	43.4	3.7	3294	1	US-08-409-995-1
17	43.4	3.7	3294	3	US-08-685-467-1
18	43.4	3.7	3294	3	US-08-913-942-1
19	43.4	3.7	3294	4	US-09-268-347-43
20	43.4	3.7	3300	4	US-09-268-347-31
21	43.4	3.7	7253	4	US-09-268-347-35
22	43.4	3.7	7291	3	US-08-913-942-3
23	43.2	3.7	3477	4	US-09-462-720-3
C 24	43.2	3.7	6243	2	US-09-056-075-1
C 25	42.8	3.7	1311	4	US-09-489-039A-4614
26	42.6	3.6	3030	4	US-09-268-347-29
27	42.6	3.6	3036	4	US-09-268-347-23

28	42.4	3.6	10467	4	US-10-204-708-2	Sequence 2, Appli
29	42.2	3.6	6866	4	US-10-204-708-20	Sequence 20, Appl
30	42.2	3.6	19233	4	US-10-204-708-45	Sequence 45, Appl
31	42	3.6	1122	4	US-09-543-681A-3292	Sequence 3292, Ap
32	41.8	3.6	837	3	US-08-998-416-288	Sequence 288, App
C 33	41.6	3.6	11049	4	US-10-204-708-23	Sequence 23, Appl
34	41.4	3.5	5562	4	US-10-204-708-63	Sequence 63, Appl
C 35	41.4	3.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
36	41	3.5	1575	4	US-09-543-681A-615	Sequence 615, App
37	41	3.5	1812	4	US-09-268-347-37	Sequence 37, Appl
C 38	40.6	3.5	3033	4	US-09-134-001C-2341	Sequence 2341, Ap
39	40.6	3.5	636	3	US-08-998-416-1137	Sequence 1137, Ap
40	40.4	3.4	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 41	40.2	3.4	14066	4	US-09-601-198-56	Sequence 56, Appl
42	40	3.4	5852	1	US-07-867-106-2	Sequence 2, Appli
43	40	3.4	6306	4	US-10-204-708-49	Sequence 49, Appl
C 44	39.8	3.4	615	3	US-08-998-416-186	Sequence 186, App
C 45	39.8	3.4	636	3	US-08-998-416-1137	Sequence 1137, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-Fls
US-08-232-463-14

Query Match 4.9%; Score 57.6; DB 1; Length 7218;

Best Local Similarity 15.6%; Pred. No. 0.00019;
Matches 81; Conservative 199; Mismatches 238; Indels 0; Gaps 0;

Qy 556 TAAATATTACTAGACCTGGGTACTTTATAGATGATCTTTATCGTATGGAACAAA 615
Db 876 TAAAGCATTCATACACACAGCTAGTTTACCAACATTCAGATAAGTAATAGAT 935
Qy 616 TACATAATATCAATAGTTGTCTAAAGAATTGCAAACTGGTTAGCCAACTACAGC 675
Db 936 TCAAAATATTATAACCGTTTACGTTGAAATGTCCTAGTCGAGTCGCTACTATACTA 995
Qy 676 ATTGCTATGTTAGTGCACCAATAGGTGTAGGCAAAAGAGCGCTTCTGTCGGTAGG 735
Db 996 TTTTTCCTTCGTTGGCCATAGCTCACAGAATTAAATCCGAGCTTGGCTGAGGTCGAG 1055
Qy 736 AGGTTATAGAGATAAACTGCAATAGCATTTGGTGTGGCTCACGCACTACTGATCGCTT 795
Db 1056 GAGCTGCGATYY 1115
Qy 796 TACGCTAAAGCGGTGTAGCGTTCAATACCTACAATGCGCGCATGCTTTATGCTGCTC 855
Db 1116 YY 1175
Qy 856 TGTGTTATGAATCTAATCAATCACTAGTTTAATCACTAATCGTTTGTGTTATAATAAAA 915
Db 1176 YY 1235
Qy 916 GGCTAAATGTTCTCCTCACATTTAGCCTTCTTATTATCTTTGTTATAGCTTTTCTG 975
Db 1236 YY 1295
Qy 976 TTATAAAACCGTTTTTAGCCACTTTTATTAAATAGCTTTTAAAGCCTTATCAATCAGTT 1035
Db 1296 YY 1355
Qy 1036 CTACTTTCACCTTTTTCACATATATATCCGCCACTTCT 1073
Db 1356 YY 1393

RESULT 2
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 4.6%; Score 54; DB 4; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 0.0066;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCAGTTTACCTTTTCCACTTTTTCACCATATATCGCCACTTCTTAAACGGTAA 1083
Db 115616 ATTAAAGTTTGGACACTTTCGCTGTTTATCATATATCGCCACTTCAAGAATAGTGA 115557
Qy 1084 TATTAAGTTGTTAGCCTAAATTTGGTACCTTCTATCGGAAATTTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTTCCAAATCGTCTTCATCGGGTATTTCTCTTAAGTTTCTA 115497
Qy 1144 AAATAAGCGTTAAAGTGCG 1165
Db 115496 AGATTAAACCATTAAGTAGC 115475

RESULT 3
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:


```
;
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          4.6%; Score 54; DB 4; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 0.0066;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1024 ATTCAATCAGTCTTACCTTTTTCACCTTTTTCACCAATATTATCGGCACCTTCTAAACGGTAA 1083
Db 115616 ATTAAGTTTGACGACTTTCGCTGTTTATCATATTATCGCCACCTTCAAGATAGTA 115557

QY 1084 TATTAAGTTGGTTAGCCTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTTCTA 1143
Db 115556 TTAGCAACCAATCAATTTGCAAAATCGTCTTCATCGGTAATTTCTCTAAGTGTCTA 115497

QY 1144 AATTAAGCGTTAAAGTGGG 1165
Db 115496 AGATTAAACCAATTAAGTAGC 115475

RESULT 4
US-09-336-447A-4
; Sequence 4, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2596
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-4

Query Match          4.6%; Score 53.2; DB 4; Length 2596;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGTCTAT 684
Db 2111 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCTGCCAAGCTGCCCTAAGTGG 2170

QY 685 GTTAGTGCACCAAAATGGTTAGCAAAAACAGCGTTTCTGCTCGGTAGAGGTTATAG 744
Db 2171 TCTATTCCAGCTTTATAGCGTTGGTAAGTTTAATGCGACCGCTGCACCTTGGTGCTATGG 2230

QY 745 AGATAAACTGCATTAGCCATTGGTTCGGCTACGCAATTAATGATCGCTTTACCGCTAA 804
Db 2231 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTATCGTGAATCCAAATCTGCGGTTAA 2290

QY 805 AGCGGGTGTAGCGTTCAATACC 826
Db 2291 AGCTGGTGGCGGTAATTAATACC 2312

RESULT 5
US-09-540-236-756
; Sequence 756, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 756
; LENGTH: 2604
; TYPE: DNA
; ORGANISM: M. catarrhalis
; US-09-540-236-756

Query Match          4.6%; Score 53.2; DB 4; Length 2604;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGTCTAT 684
Db 2352 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCTGCCAAGCTGCCCTAAGTGG 2411

QY 685 GTTAGTGCACCAAAATGGTTAGCAAAAACAGCGTTTCTGCTCGGTAGAGGTTATAG 744
Db 2412 TCTATTCCAGCTTTATAGCGTTTAAAGTTTAATGCGACCGCTGCACCTTGGTGCTATGG 2471

QY 745 AGATAAACTGCATTAGCCATTGGTTCGGCTCAGCAATTAATGATCGCTTTACCGCTAA 804
Db 2472 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTGAATCCAAATCTGCGGTTAA 2531

QY 805 AGCGGGTGTAGCGTTCAATACC 826
Db 2532 AGCTGGTGGCGGTAATTAATACC 2553

RESULT 6
US-09-336-447A-12
; Sequence 12, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 2673
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-12

Query Match          4.6%; Score 53.2; DB 4; Length 2673;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGTCTAT 684
Db 2180 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCTGCCAAGCTGCCCTAAGTGG 2239

QY 685 GTTAGTGCACCAAAATGGTTAGCAAAAACAGCGTTTCTGCTCGGTAGAGGTTATAG 744
Db 2240 TCTATTCCAGCTTTATAGCGTTTAAAGTTTAATGCGACCGCTGCACCTTGGTGCTATGG 2299

QY 745 AGATAAACTGCATTAGCCATTGGTTCGGCTCAGCAATTAATGATCGCTTTACCGCTAA 804
Db 2300 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTGAATCCAAATCTGCGGTTAA 2359
```

```
Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 2360 AGCTGGTGGCGGATTAATACC 2381

RESULT 7
US-09-336-447A-16
; Sequence 16, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 4228
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-16

Query Match 4.6%; Score 53.2; DB 4; Length 4228;
Best Local Similarity 54.0%; Pred. No. 0.0018;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGCTAAAGATTGCAAACTGTTTAGCCAAACCAATCAGCATTTGCTAT 684
Db 3629 TATCAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCTGCCAAGTCGCCCTAAGTGG 3688

Qy 685 GTTAGTCAACCAATGGTGTAGGCAAAACAGAGGTTTCTGCTCGGTAGGAGTTATAG 744
Db 3689 TCTATTCCAGCTTTATAGCGTTGGTAAAGTTTAAATGCGACCGCTGCACTTGGTGGCTAAG 3748

Qy 745 AGATAAACTGCATTAGCATTTGTCGGCTACGCATTAATCTGATCGCTTTACCGCTAA 804
Db 3749 CTCAAAATCTCGGTTGCTATCGGTGCTGCTATCGTGAATCCAAATCTGGCGGTTAA 3808

Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 3809 AGCTGGTGGCGGATTAATACC 3830

RESULT 8
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; APPLICANT: BEIG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596.002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:

US-09-596-002-38
Query Match 4.6%; Score 53.2; DB 4; Length 94750;
Best Local Similarity 54.0%; Pred. No. 0.0044;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGCTAAAGATTGCAAACTGTTTAGCCAAACCAATCAGCATTTGCTAT 684
Db 26785 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCTGCCAAGCTGCCCTAAGTGG 26844

Qy 685 GTTAGTCAACCAATGGTGTAGGCAAAACAGAGGTTTCTGCTCGGTAGGAGTTATAG 744
Db 26845 TCTATTCCAGCTTTATAGCGTTGGTAAAGTTTAAATGCGACCGCTGCACTTGGTGGCTATGG 26904

Qy 745 AGATAAACTGCATTAGCATTTGTCGGCTACGCATTAATCTGATCGCTTTACCGCTAA 804
Db 26905 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTGTGAATCCAAATCTGGCGGTTAA 26964

Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 26965 AGCTGGTGGCGGATTAATACC 26986

RESULT 9
US-09-336-447A-8
; Sequence 8, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3295
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-8

Query Match 4.5%; Score 52.4; DB 4; Length 3295;
Best Local Similarity 55.5%; Pred. No. 0.0026;
Matches 101; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 645 AATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTCAACCAATAGTG 704
Db 2854 AGTTGAAAACGGTATGCTGCTGCCAAGCTGCCCTAAGTGGTCTATTCAGGCTTATAGCG 2913

Qy 705 TAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAAAACCTGCATTAGCCA 764
Db 2914 TTGGTAAGTTTAAATGCGACCGCTGCACTTGGTGGCTATGGCTCAAAATCTCGGTTGCTA 2973

Qy 765 TTGGTGTGCGCTCAGCATTAATCTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATA 824
Db 2974 TCGGTGCTGCTATCGTGTGAATCCAAATCTGGCGGTTTAAAGCTGGTGGCGGATTAATA 3033

Qy 825 CC 826
Db 3034 CC 3035

RESULT 10
US-10-204-708-37
; Sequence 37, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
```

```

; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      3.9%; Score 45.4; DB 4; Length 640681;
Best Local Similarity 48.0%; Pred. No. 0.53;
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 458 AATTGTATAGTTATTTTAAACGATTTTAAGACACGATTTTAAATTAATAAGTTCTTGATGCA 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209572 AATTTTAAATATATAGGACACCTGATAAATTTTATTAAATTTAAAGAACTAGACGTA 209513

Qy 518 CGTATTTCCAAAATAAACAAAATATTCATACTATAAGTAATAATTTACTAGAACTGGT 577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209512 TATTTTTCAAATAAATAACAAACAGCGCGCTTAATGTAGTAAATGATTATGATGCA 209453

Qy 578 ACTTATTTAGATGATTCCTATCGTATGATGAACAAAATAACATAATATCAATAAGTTG 637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209452 ACAATAGACAAAAGAACAATTTTATAAACCCTTGAAGAAGTTAATAATTAATATCCA 209333

Qy 638 TCTAAAGAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTATGTTAGTGCAACCA 697
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209392 TTATACAAAGTAAAAAATGGTGAATGCACTAGCATTTTATACATAATTTATATCAACCA 209333

Qy 698 AATGGTGTAGGCAAAACAGAGGTTTCTGCTG 728
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209332 ATTCCTCTGGTACTAAGAGGCTACGGTCTG 209302

RESULT 12
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      3.9%; Score 45; DB 4; Length 640681;
Best Local Similarity 47.8%; Pred. No. 0.66;
Matches 162; Conservative 0; Mismatches 175; Indels 2; Gaps 1;

Qy 422 GATCACTTCGAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAACGAT 481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165491 GATGCCATCTGAAGAGTTCAAAAAGAAGCAAAAGAATTAATATATCAATCTTTCAAC 165508

Qy 482 TTAAGACACGATTTTAAATTAATAAGTTCTTCGATGCACGTATTTCCAAAATAAACAATAAT 541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165551 TTCAGAAAATTTT--ATAATAATATTAATAATATAAATTTCTTTAATATTACTGATT 165608

Qy 542 ATTGATACTATAAGTAATAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTTTATCGT 601
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165609 TTTTAGATGATAATATTATTATATATTTTAAACTGAGTGGTGAGAGATATATGATTGATTTA 165668

Qy 602 ATGATGAACAAAATACACATAATATCAATAGTTGTCTAAAGATTCGAAGATTCGAACCTGTTTA 661
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165669 ATGCCAAAAAATAATACATTTTATTAAAAAATAATAAGAAAGACTCAAAATAATTCA 165728

```

QY 662 GCCAACCAATCAGCATTTCTAATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTT 721
Db 165729 GTAATTAATAATTTATTTGATTTCTTTGGTAAAGAATTTGTTTTCATCAAAATTACCTTA 165788
QY 722 TCTGTCGGGTAGGAGTTATAGAGATAAAACATGCATTA 760
Db 165789 ACCGGTTTTCTTATTAATTTGGATTGATAAGACATTATTA 165827

RESULT 13

US-10-204-708-22
; Sequence 22, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 22
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-22

Query Match 3.8%; Score 44; DB 4; Length 11049;

Best Local Similarity 51.5%; Pred. No. 0.36;

Matches 101; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 822 ATACCTACATGCGGCATGCTTATGGTCTTCTGTTAGTGCATTAATCTAATCATATAC 881
Db 1402 ATGATTAATAATGAATTTTATGTTTTTATTTATTTAGTCTGTTATGTTTATTTAGTGA 1461
QY 882 GTTTAATCACTAATCGTTTGGTTATAATAAAAGGCTAAATGTTCTCTCACAATTTAG 941
Db 1462 ATGTTTTTGTATTTATTTAGTTAGTAGAATGTTGGGTATTTATTTATTTATTTATTTAG 1521
QY 942 CCTTCTTATTTATTTCTTTGTTATAGCTTTTGTGTTATATAAACCGTTTTTTAGCCACTTT 1001
Db 1522 GTTTATTAATTTTATTTTAAATGTTTATTTGAAGGTATATATTAGATTTTATTAACGTAT 1581
QY 1002 TATTAATTAAGCTTTT 1017
Db 1582 TTTTGTGTAATTTT 1597

RESULT 14

US-08-913-942-14
; Sequence 14, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP

; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/913,942
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RPT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2037 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-913-942-14

Query Match 3.7%; Score 43.4; DB 3; Length 2037;

Best Local Similarity 55.7%; Pred. No. 0.31;

Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 623 AATATCAATAAGTTGTCTTAAGAATTCGAAACTGGTTTAGCCAAACCAATCAGCATTTGCT 682
Db 1789 AAAGTGAATAAGTGGGCAACGTCAGATGCAGTACTGCAAGTCATTAGCGGCTTCA 1848
QY 683 ATGTTAGTGCACCAATGGTGTAGGCAAAACGACGCTTCTGCTGCGGTAGGAGGTTAT 742
Db 1849 CAGTTACCAACAGCCACTATGCCAGGTAAATCAATGGTTTCTATTGCGGGAAGTAGTTAT 1908
QY 743 AGAGATAAACTGCATTAGCCATTGGTGT 771
Db 1909 CAAGTCAAATGGTTTAGCTATCGGGT 1937

RESULT 15

US-09-268-347-25
; Sequence 25, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-09-268-347-25

Query Match 3.7%; Score 43.4; DB 4; Length 2079;

Best Local Similarity 55.7%; Pred. No. 0.31;
 Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy	623	AAATATCAATTAAGTTGCTTAAGAATTCGAAACCTGGTTTAGCCAAACCAATCAGCATTTGCT	682
Db	1820	AAAGTGAATAAAGTGGGCAACCGTCAGATGCAGTACTGCAAGTGCATTAGCGGCTTCA	1879
Qy	683	ATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGGTTAT	742
Db	1880	CAGTTACCAAGCCACTATGCCAGGTAATCAATGGTTTCTATTGCGGGAAGTAGTTAT	1939
Qy	743	AGAGATAAAACTGCATTAGCCATTGGTGT	771
Db	1940	CAAGTCAAAATGTTTAGCTATCGGGT	1968

Search completed: May 15, 2004, 05:23:09
 Job time : 128 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 01:31:40 ; Search time 561 Seconds
(without alignments)
8844.733 Million cell updates/sec

Title: US-10-030-529a-1
Perfect score: 1168
Sequence: 1 ataaacacgtcattgacatt.....aagccgttaagtgcgac 1168

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	1168	4	Aaf25262 Nucleotid
2	1151	98.5	1189	4	Aaf25270 Nucleotid
3	1139	97.5	1197	4	Aaf25266 Nucleotid
4	1092.6	93.5	1205	4	Aaf25263 Nucleotid
5	1090	93.3	1231	4	Aaf25268 Nucleotid
6	866.8	74.2	1047	4	Aaf25269 Nucleotid
7	839.6	71.9	952	4	Aaf25264 Nucleotid
8	829.2	71.0	899	4	Aaf25265 Nucleotid
9	818	70.0	923	4	Aaf25267 Nucleotid
10	67.2	5.8	4426	3	Aac79587 Virulence
11	67.2	5.8	4426	6	Abq83464 Pasteurel
12	59.8	5.1	1095	3	Aaz54498 Neisseria
13	59.8	5.1	16526	3	Aaa81472 N. mening
14	59.8	5.1	110000	3	Continuation (7 of
15	59.8	5.1	172325	3	Aaf21613 Neisseria
16	59.8	5.1	349980	3	Aaf21612 Neisseria
17	59.2	5.1	972	7	Abx93390 Neisseria
18	57.6	4.9	1938	4	Aad17051 N. mening
19	57.6	4.9	1938	5	Aas43898 Neisseria
20	57.6	4.9	1947	4	Aad17043 Neisseria
21	57.6	4.9	1947	5	Aas43878 Neisseria
22	57.6	4.9	2388	4	Aad17035 N. mening
23	57.6	4.9	2388	5	Aas43870 Neisseria

24	57.6	4.9	2421	4	AAD17050
25	57.6	4.9	2421	4	AAD17048
26	57.6	4.9	2421	5	AAS43897
27	57.6	4.9	2421	5	AAS43895
28	57.6	4.9	2583	4	AAD17038
29	57.6	4.9	2583	5	AAS43873
30	57.6	4.9	3798	4	AAD17041
31	57.6	4.9	4335	4	AAD17052
32	57.6	4.9	4335	5	AAS43899
33	57.6	4.9	4344	5	AAS43876
34	57.2	4.9	8139	7	ACA43016
35	54	4.6	2000	7	ADA71938
36	54	4.6	110000	2	AT42063_01
37	53.2	4.6	2596	2	AAV41342
38	53.2	4.6	2673	2	AAV41346
39	53.2	4.6	4228	2	AAV41348
40	53.2	4.6	94750	4	AAF28551
41	52.4	4.5	3295	2	AAV41344
42	52.2	4.5	558	7	AAD50226
43	51.8	4.4	18624	6	ABL333702
44	50.8	4.3	9117	6	ABL33068
45	49	4.2	18683	6	ABL32313

ALIGNMENTS

RESULT 1
AAF25262
ID AAF25262 standard; DNA; 1168 BP.
XX
AC AAF25262;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA locus.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease; ss.
XX
OS Haemophilus ducreyi.
FH Key Location/Qualifiers
FT CDS 101..874
FT /*tag= a
FT /product= "DsrA protein"
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
PI Elkins C;
PI WPI; 2001-138311/14.
DR P-PSDB; AAB31705.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Fig 3; 80pp; English.
XX
CC The present sequence encodes a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA

CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 1168 BP; 357 A; 190 C; 209 G; 412 T; 0 U; 0 Other;

Query Match 100.0%; Score 1168; DB 4; Length 1168;
 Best Local Similarity 100.0%; Pred. No. 1.1e-243;
 Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATAGCTCATTCGACATTTTAAATGTAAGGTAGATAAGAAAGTAAATTCATATT 60
 Db 1 ATAAATAGCTCATTCGACATTTTAAATGTAAGGTAGATAAGAAAGTAAATTCATATT 60
 Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTAGT 120
 Db 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTAGT 120
 Qy 121 TCGCGTAGTGGGATTAGCTTCTTACTATTACAAATGGCTCAGCAGCGCCCAAGTT 180
 Db 121 TCGCGTAGTGGGATTAGCTTCTTACTATTACAAATGGCTCAGCAGCGCCCAAGTT 180
 Qy 181 TGCTGGAGTATCTTTTGTAATGATGATGATGATGATGATGATGATGATGATGATG 240
 Db 181 TGCTGGAGTATCTTTTGTAATGATGATGATGATGATGATGATGATGATGATGATG 240
 Qy 241 GTCTAAATCAAGCGGTTTCGATATTAAAGTCCAGGGATTAATAATGAAGCCAAAGATG 300
 Db 241 GTCTAAATCAAGCGGTTTCGATATTAAAGTCCAGGGATTAATAATGAAGCCAAAGATG 300
 Qy 301 GATTCTTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTCT 360
 Db 301 GATTCTTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTCT 360
 Qy 361 CGTGACATATGCTCTGCGGTTCTCTAGCCCTATATGTTATATGCTGATGCTGATCC 420
 Db 361 CGTGACATATGCTCTGCGGTTCTCTAGCCCTATATGTTATATGCTGATGCTGATCC 420
 Qy 421 TGATCACTTGGATAAATCGGACGCTGAAATGAAATTTGATGATGATGATGATGATG 480
 Db 421 TGATCACTTGGATAAATCGGACGCTGAAATGAAATTTGATGATGATGATGATGATG 480
 Qy 481 TTTAAGCACGATTTTAAATTTAAAGTTCTTGTGATGACGATTTTCCAAAATAAACAAA 540
 Db 481 TTTAAGCACGATTTTAAATTTAAAGTTCTTGTGATGACGATTTTCCAAAATAAACAAA 540
 Qy 541 TATTGATCTATATAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATGATGATG 600
 Db 541 TATTGATCTATATAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATGATGATG 600
 Qy 601 TATGATGCAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTCGAAATCTGTTT 660
 Db 601 TATGATGCAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTCGAAATCTGTTT 660
 Qy 661 AGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
 Db 661 AGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
 Qy 721 TTCTGCTCGGTAGAGGTTTATAGAGATAAAATGCAATTAGCCATTTGGTGTGCGGTCAAG 780
 Db 721 TTCTGCTCGGTAGAGGTTTATAGAGATAAAATGCAATTAGCCATTTGGTGTGCGGTCAAG 780
 Qy 781 CATTACTGATCGTTTACCGCTTAAAGCGGTTAGCGTTTCAATACCTACATGCGGCAT 840
 Db 781 CATTACTGATCGTTTACCGCTTAAAGCGGTTAGCGTTTCAATACCTACATGCGGCAT 840
 Qy 841 GTCTTATGGTCTTCTGTGTTTATGAATTTCTAATCATTAGCTTTAATCACTAATCGTTT 900

Db 841 GTCTTATGGTCTTCTGTGTTTATGAATTTCAATCATAGTTTAATCACTAATCGTTT 900
 Qy 901 TGGTTATAATAAAGCGCTAAATGCTTCTCTCCTCAGCATTTAGCCCTTCTTATTTATCTTTG 960
 Db 901 TGGTTATAATAAAGCGCTAAATGCTTCTCTCCTCAGCATTTAGCCCTTCTTATTTATCTTTG 960
 Qy 961 TTATAGCTTTTGTGTTTATAAACCCTGTTTATAGCCACTTTTATTAATTAAGCTTTTAAAG 1020
 Db 961 TTATAGCTTTTGTGTTTATAAACCCTGTTTATAGCCACTTTTATTAATTAAGCTTTTAAAG 1020
 Qy 1021 CCTATTCAATCAGTTCTACTTTTCACTTTTTCACCATTTATTCGCCACTTCTCTAAAACGG 1080
 Db 1021 CCTATTCAATCAGTTCTACTTTTCACTTTTTCACCATTTATTCGCCACTTCTCTAAAACGG 1080
 Qy 1081 TAATATTAAGTTGGTTTATAGCTTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTT 1140
 Db 1081 TAATATTAAGTTGGTTTATAGCTTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTT 1140
 Qy 1141 CTAAATTAAGCCGTTTAAAGGTGCGGAC 1168
 Db 1141 CTAAATTAAGCCGTTTAAAGGTGCGGAC 1168

RESULT 2
 AAF25270
 ID AAF25270 standard; DNA; 1189 BP.
 XX AAF25270;
 AC AAF25270;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Nucleotide sequence of the DsrA protein from strain OP406.
 XX
 KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease; ss.
 XX
 OS Haemophilus ducreyi.
 XX
 PH Key Location/Qualifiers
 FT CDS 139..912
 FT /*tag= a
 FT /product= "DsrA protein"
 XX
 XX W0200104138-A1.
 XX
 XX 18-JAN-2001.
 XX
 XX 07-JUL-2000; 2000WO-US018834.
 XX
 XX 09-JUL-1999; 99US-0143257P.
 XX
 XX (UNCL-) UNIV NORTH CAROLINA.
 XX (ELKI/) ELKINS C.
 XX
 XX Elkins C;
 XX
 XX WPI; 2001-138311/14.
 XX P-PSDB; AAB31713.
 XX

Novel purified DsrA protein of Haemophilus ducreyi which confers serum resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.

Claim 1; Page 58; 80pp; English.

The present sequence encodes a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX

SQ Sequence 1197 BP; 369 A; 191 C; 204 G; 433 T; 0 U; 0 Other;

Query Match 97.5%; Score 1139; DB 4; Length 1197;

Best Local Similarity 99.9%; Pred. No. 2.1e-237;

Matches 1150; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ATAAATAGCATTGACATTTTAAATGTAAGTAGAATAAGAAAGTAAATCTATATT 60

Db 47 ATAAATAGCATTGACATTTTAAATGTAAGTAGAATAAGAAAGTAAATCTATATT 105

Qy 61 TACAATCAAGATTGACAAATTTTAAATGTAAGTAGAATAAGAAAGTAAATCTATATT 120

Db 107 TACAATCAAGATTGACAAATTTTAAATGTAAGTAGAATAAGAAAGTAAATCTATATT 166

Qy 121 TGCCGTAGTGGGATTAGCTTCTTACTATTACAAATGCTCAGCAGCGCCCAAGTT 180

Db 167 TGCCGTAGTGGGATTAGCTTCTTACTATTACAAATGCTCAGCAGCGCCCAAGTT 225

Qy 181 TGCTGGAGTATCTTCTTTGTATAGCTATGATATGATATGTAAGGGTAAATGGACTTG 240

Db 227 TGCTGGAGTATCTTCTTTGTATAGCTATGATATGATATGTAAGGGTAAATGGACTTG 285

Qy 241 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGGAATTAAGTGAAGGTAAGGTAAGT 300

Db 287 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGGAATTAAGTGAAGGTAAGGTAAGT 346

Qy 301 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTTATGCTTATATCTCTGTCT 360

Db 347 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTTATGCTTATATCTCTGTCT 406

Qy 361 CGTGACATATGCTCCTGGCGTTTCTCTAGCCCTATATGTTATATCGATGTCGTATCC 420

Db 407 CGTGACATATGCTCCTGGCGTTTCTCTAGCCCTATATGTTATATCGATGTCGTATCC 465

Qy 421 TGATCAACTTGGATAAATCGGCAGCAGCTGAAATGAAATTTGTATAGTTATTTAAACA 480

Db 467 TGATCAACTTGGATAAATCGGCAGCAGCTGAAATGAAATTTGTATAGTTATTTAAACA 526

Qy 481 TTTAAGCACGATTTTAAATTTAAAGTTCTTGTAGCAGCTATTTCCAAAATAAACAAA 540

Db 527 TTTAAGCACGATTTTAAATTTAAAGTTCTTGTAGCAGCTATTTCCAAAATAAACAAA 586

Qy 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGTTCTTATCG 600

Db 587 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGTTCTTATCG 646

Qy 601 TATGATGGAACAAATACACATAATATCAATAAGTTGTCTAAAGAAATTCGAACTGGTTT 660

Db 647 TATGATGGAACAAATACACATAATATCAATAAGTTGTCTAAAGAAATTCGAACTGGTTT 706

Qy 661 AGCCAACCAATCAGCATTTGTATGTTAGTGCACCAATGGTGTAGGCAAAACGAGCGT 720

Db 707 AGCCAACCAATCAGCATTTGTATGTTAGTGCACCAATGGTGTAGGCAAAACGAGCGT 765

Qy 721 TTCTGTCGGTAGAGGTTATAGAGTAAATTCGATTTAGCCATTTGGTGTGCGTCAAG 780

Db 767 TTCTGTCGGTAGAGGTTATAGAGTAAATTCGATTTAGCCATTTGGTGTGCGTCAAG 826

Qy 781 CATTACTGATCGCTTACCGTTAAAGCGGTGAGCGTTCAATACCTACATGCGGCAT 840

Db 827 CATTACTGATCGCTTACCGTTAAAGCGGTGAGCGTTCAATACCTACATGCGGCAT 886

Qy 841 GTCTTATGGTCTTCTGTTGTTATGAAATCTTAATCATTAGCTTTAATCACTAATCGTTT 900

Db 887 GTCTTATGGTCTTCTGTTGTTATGAAATCTTAATCATTAGCTTTAATCACTAATCGTTT 946

Qy 901 TGGTTATTAATAAAGGCTAAATGTTTCTCTCAGATTTAGCCCTTCTATTATCTTTG 960

Db 947 TGGTTATTAATAAAGGCTAAATGTTTCTCTCAGATTTAGCCCTTCTATTATCTTTG 1006

Qy 961 TTATAG-CTTTTCTGTTATATAAACCGTTTTTTTAGCCACTTTTATTATTAAGCTTTAA 1019

Db 1007 TTATAGCTTTTCTGTTATATAAACCGTTTTTTTAGCCACTTTTATTATTAAGCTTTAA 1066

Qy 1020 GCCTATTCAATCAGTCTTACTTTTACCTTTTTCACATATATCCGCACCTTCTAAACG 1079

Db 1067 GCCTATTCAATCAGTCTTACTTTTACCTTTTTCACATATATCCGCACCTTCTAAACG 1126

Qy 1080 GTAATATTAGTGGTTTACGCTTAAATTTGGGTACCTTCTATCGAAATTTTCTCTAAATGT 1139

Db 1127 GTAATATTAGTGGTTTACGCTTAAATTTGGGTACCTTCTATCGAAATTTTCTCTAAATGT 1186

Qy 1140 TCTAAAAATTAA 1150

Db 1187 TCTAAAAATTAA 1197

RESULT 4

AAF25263

ID AAF25263 standard; DNA; 1205 BP.

XX AAF25263;

XX DT 30-APR-2001 (first entry)

XX Nucleotide sequence of the DsrA protein from strain CIPA75.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;

XX chancroid disease; genital ulcer disease; ss.

XX Haemophilus ducreyi.

XX Key Location/Qualifiers

XX CDS 135..929

XX /*tag= a

XX /product= "DsrA protein"

XX W0200104138-Al.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UNCL-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum

XX resistance to the bacteria used to produce vaccines that induce immune

XX response against the bacteria in subject at risk of developing chancroid.

XX Claim 1; Page 53; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus

XX ducreyi. DsrA is an outer membrane protein that confers serum resistance

XX to the bacteria. DsrA antisense oligonucleotides are useful for detecting

XX a polynucleotide which encodes DsrA in a biological sample. The DsrA

XX polypeptide is used to produce vaccine compositions, which are useful for

XX inducing a protective immune response in a subject at risk of developing

XX Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

XX transmitted by sexual contact. DsrA, its catalytic or immunogenic

XX fragments or peptides can be used for screening libraries of compounds in

XX a variety of drug screening techniques. The proteins and peptides may be

XX used as antigens in immunoassays for detection of Haemophilus ducreyi in

XX various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .

XX The nucleic acids are useful for the preparation of DsrA proteins

XX

SQ Sequence 1205 BP; 360 A; 190 C; 200 G; 435 T; 0 U; 0 Other;

Query Match 93.5%; Score 1092.6; DB 4; Length 1205;
Best Local Similarity 97.4%; Pred. No. 2.4e-227;
Matches 1141; Conservative 0; Mismatches 4; Indels 26; Gaps 2;

QY 1 ATAATAAGTACGATGACATTTTAAATGTAAGGTAGAAATAAGAAAGTAAATTTCTATATT 60
DB 40 ATAATAAGTACGATGACATTTTAAATGTAAGGTAGAAATAAGAAAGTAAATTTCTATATT 94

QY 61 TACAATCAAGATGACAAATTAATTTACTTAATGAGGTGATTAATGAAAAATTAAGTTTGTAGT 120
DB 95 TACAATCAAGATGACAAATTAATTTACTTAATGAGGTGATTAATGAAAAATTAAGTTTGTAGT 154

QY 121 TGGCGTAGTGGATAGCTTCTTCTACTATTACAAATGCTCAGCAGCGCCGCAAGTT 180
DB 155 TGGCGTAGTGGATAGCTTCTTCTACTATTACAAATGCTCAGCAGCGCCGCAAGTT 214

QY 181 TGGTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTAAATGGACTTG 240
DB 215 TGGTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTAAATGGACTTG 274

QY 241 GTCTAATCAAGCGGTTTCGATATTAAAGTSCCAGGGATTAATAATGAAGCCAAAGAAATG 300
DB 275 GTCTAATCAAGCGGTTTCGATATTAAAGTSCCAGGGATTAATAATGAAGCCAAAGAAATG 334

QY 301 GATTTCTAAGCAGCTACTTATCTTGAATACACATTAATGCTTATATCTCTGTTCT 360
DB 335 GATTTCTAAGCAGCTACTTATCTTGAATACACATTAATGCTTATATCTCTGTTCT 394

QY 361 CGTGACATATGCTCCTCGCGTTTCTCTAGCCCTATCTGTTATATCGATGCTGATCC 420
DB 395 CGTGACATATGCTCCTCGCGTTTCTCTAGCTCTATCTGTTATATCGATGCTGATCC 454

QY 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 480
DB 455 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 514

QY 481 TTTAAGCAGCATTTTAAATTTAAAGTTCTGTAGTCAGGATTTTCCAAAAATAACAAAA 540
DB 515 TTTAAGCAGCATTTTAAATTTAAAGTTCTGTAGTCAGGATTTTCCAAAAATAACAAAA 574

QY 541 TATTGATCTATAGTAAATTTTACTAGACCTGGGTACTTATTTAGATGATCTTATCG 600
DB 575 TATTGATCTATAGTAAATTTTACTAGACCTGGGTACTTATTTAGATGATCTTATCG 634

QY 601 TATGATGGAAC-----AAAATACACATATATCAATAAATAACATATATCAATAAGTTGTC 639
DB 635 TATGATGGAACAAATACACATATATCAATAAATAACATATATCAATAAGTTGTC 694

QY 640 TAAAGAAATGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTATGTAGTGCACCAAA 699
DB 695 TAAAGAAATGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTATGTAGTGCACCAAA 754

QY 700 TGGTGTAGGCAAAACGCGGTTCTGCTGCGGTAGGAGTTATAGATAAACCTGCATT 759
DB 755 TGGTGTAGGCAAAACGCGGTTCTGCTGCGGTAGGAGTTATAGATAAACCTGCATT 814

QY 760 AGCCATTGGTGTGGCTCAGCATTTACTGATGCTTTTACCGCTAAAGCGGTTGAGCGTT 819
DB 815 AGCCATTGGTGTGGCTCAGCATTTACTGATGCTTTTACCGCTAAAGCGGTTGAGCGTT 874

QY 820 CAATACCTTACAAATCGCGCATGTCTTATGCGTCTTCTGTTGTTATGAATCTTAATCATT 879
DB 875 CAATACCTTACAAATCGCGCATGTCTTATGCGTCTTCTGTTGTTATGAATCTTAATCATT 934

QY 880 ACGTTTAACTACTAATCGTTTGTGTTATATAAAGGCTAAATGTTTCTCTCCTCACATTT 939
DB 935 ACGTTTAACTACTAATCGTTTGTGTTATATAAAGGCTAAATGTTTCTCTCCTCACATTT 994

QY 940 AGCCTTTCTTATTTATCTTCTGTTATAGCTTTGCTGTTTATAAACCCTTTTTTAGCCACT 999
DB 995 AGCCTTTCTTATTTATCTTCTGTTATAGCTTTGCTGTTTATAAACCCTTTTTTAGCCACT 1054

QY 1000 TTTTAAATTAAGCTTTTAAGCCTATTCAATCAGTTCTACTTTTCACTTTTCCACCATAT 1059
DB 1055 TTTTAAATTAAGCTTTTAAGCCTATTCAATCAGTTCTACTTTTCACTTTTCCACCATAT 1114
QY 1060 TATCCCGCAGCTTCTAAAAACGGTAATATTAAAGTTGGTTTACCTTAAATTTGGGTACCTTTCTA 1119
DB 1115 TATCCCGCAGCTTCTAAAAACGGTAATATTAAAGTTGGTTTACCTTAAATTTGGGTACCTTTCTA 1174

QY 1120 TCGGAATTTTCTTAAATGTTCTAAAAATTAA 1150
DB 1175 TCGGAATTTTCTTAAATGTTCTAAAAATTAA 1205

RESULT 5
AAF25268 standard; DNA; 1231 BP.

XX AAF25268;
XX AC
XX 30-APR-2001 (first entry)

XX Nucleotide sequence of the DsrA protein from strain V-1157.
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease; ss.
XX Haemophilus ducreyi.

XX Key Location/Qualifiers
XX CDS 140..955
XX /*tag= a
XX /product= "DsrA protein"

XX WO200104138-A1.
XX PN
XX 18-JAN-2001.
XX PD
XX 07-JUL-2000; 2000WO-US018834.
XX PF
XX 09-JUL-1999; 99US-0143257P.
XX PR
XX (UVNC-) UNIV NORTH CAROLINA.
XX (ELKI/) ELKINS C.
XX PA
XX Elkins C;
XX PI
XX WPI; 2001-138311/14.
XX P-PSDB; AAB31711.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.

XX Claim 1; Page 56-57; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of Haemophilus ducreyi in various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .

XX The nucleic acids are useful for the preparation of DsrA proteins

SQ Sequence 1231 BP; 391 A; 196 C; 202 G; 442 T; 0 U; 0 Other;

Query Match 93.3%; Score 1090; DB 4; Length 1231;

Best Local Similarity 96.1%; Pred. No. 8.8e-227;		Matches 1145; Conservative 0; Mismatches 5; Indels 42; Gaps 1;	
Qy	1	ATAAATAGCTCATTTGACATATTTTAAATGTAAGGTAGAAATGAAGAAGTAAATTTCTATATT	60
Db	40	ATAAATAGCTCATTTGACATATTTTAAATGTAAGGTAGAAATGAAGAAGTAAATTTCTATATT	99
Qy	61	TACAATCAAGATTGACAAATTTATTTACTTTAATGAGGTGATATATGAAAAATTAATGTTTAGT	120
Db	100	TACAATCAAGATTGACAAATTTATTTACTTTAATGAGGTGATATATGAAAAATTAATGTTTAGT	159
Qy	121	TGCCGTAGTGGGATAGCTTGTCTTACTATTTACAAATGGCTCAGACGCCGCAAGTT	180
Db	160	TGCCGTAGTGGGATAGCTTGTCTTACTATTTACAAATGGCTCAGACGCCGCAAGTT	219
Qy	181	TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTAAGTAAATGGACTTG	240
Db	220	TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTAAGTAAATGGACTTG	279
Qy	241	GTCTAATGAAGCGGTTTCGATATTTAAAGTGCCAGGATTTAAATGAAGCCAAAGAAATG	300
Db	280	GTCTAATGAAGCGGTTTCGATATTTAAAGTGCCAGGATTTAAATGAAGCCAAAGAAATG	339
Qy	301	GATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATTTATGCTTTATATCTCTGTCT	360
Db	340	GATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATTTATGCTTTATATCTCTGTCT	399
Qy	361	CGTGACATATCTCTCTGCGTTCTCTAGCCCTATCTGTATATCGATGTCGTATCC	420
Db	400	CGTGACATCTCTCTGACGTTCTCTAGCTCTATCTGTATATCGATGTCGTATCC	459
Qy	421	TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGATATGTTATTTTAAACGA	480
Db	460	TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGATATGTTATTTTAAACGA	519
Qy	481	TTTAAGACACGATTTTAAATTTAAAGTCTTGATGTCAGGATTTTCCAAAAATAACAAA	540
Db	520	TTTAAGACACGATTTTAAATTTAAAGTCTTGATGTCAGGATTTTCCAAAAATAACAAA	579
Qy	541	TATTGATCTATAAGTAAATTTTACTAGACTGGGTACTTTTATAGATGTTCTTATCG	600
Db	580	TATTGATCTATAAGTAAATTTTACTAGAACTGGGTACTTTTATAGATGTTCTTATCG	639
Qy	601	TATGATGGAAC-----AAAATAC	618
Db	640	TATGATGGAACAAAATACACATAATCAATAAAAAATACACATAATCAATAAAAAATAC	699
Qy	619	ACATAATATCAATAGTTGCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAGATT	678
Db	700	ACATAATATCAATAGTTGCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAGATT	759
Qy	679	GTCTATGTTAGTCAACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTGCGTAGGAGG	738
Db	760	GTCTATGTTAGTCAACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTGCGTAGGAGG	819
Qy	739	TTATAGAGATAAACTGCATTAGCCATTGTTGTCGGCTCAGCATTTACTGATCGCTTAC	798
Db	820	TTATAGAGATAAACTGCATTAGCCATTGTTGTCGGCTCAGCATTTACTGATCGCTTAC	879
Qy	799	CGCTAAAGCGGGTGTAGCGTTCAATACCTACAATGCGGCGATGTTATGCGTCTTCTGT	858
Db	880	CGCTAAAGCGGGTGTAGCGTTCAATACCTACAATGCGGCGATGTTATGCGTCTTCTGT	939
Qy	859	TGGTTATCAATCTCAATCATTACGTTTAAATCACTAATCGTTTGGTTATATAAAGGC	918
Db	940	TGGTTATCAATCTCAATCATTACGTTTAAATCACTAATCGTTTGGTTATATAAAGGC	999
Qy	919	TAAATGTTTCCTCCACATTTAGCCTTCTTATTTATCTTTGTTATAGCTTTTGTGTTA	978
Db	1000	TAAATGTTTCCTCCACATTTAGCCTTCTTATTTATCTTTGTTATAGCTTTTGTGTTA	1059
Qy	979	TAAACCGTTTTTTAGCCACTTTTATTAATTAAGCTTTTAAAGCTTTTCAATCAGTTCTA	1038

Db	1060	TAAAACGGTTTTTTAGCCACTTTTATTAATAAGCTTTTAAAGCCTATTCAATCAGTTCTA	1119
Qy	1039	CTTTCACTTTTTTACCACATATTATCGCCACTTCTTAAACGGTAATATTAAAGTTGGTTTA	1098
Db	1120	CTTTCACTTTTTTACCACATATTATCGCCACTTCTTAAACGGTAATATTAAAGTTGGTTTA	1179
Qy	1099	GCTTAAATTGGGTACCTTCTATCGGAATTTTTTTCTAAATGTTCTAAAAATTAA	1150
Db	1180	GCCTAAATTGGGTACCTTCTATCGGAATTTTTTTCTAAATGTTCTAAAAATTAA	1231
RESULT 6			
AAF25269			
ID	AAF25269 standard; DNA; 1047 BP.		
XX			
AC	AAF25269;		
XX			
DT	30-APR-2001 (first entry)		
XX			
DE	Nucleotide sequence of the DsrA protein from strain M90-02.		
XX			
KW	DsrA protein; outer membrane protein; serum resistance; vaccine;		
KW	chancroid disease; genital ulcer disease; ss.		
XX	Haemophilus ducreyi.		
OS			
XX			
FH	Key Location/Qualifiers		
CDS	140..961		
FT	/*tag= a		
FT	/product= "DsrA protein"		
XX	WO200104138-A1.		
PN			
XX	18-JAN-2001.		
PD			
XX			
PF	07-JUL-2000; 2000WO-US018834.		
XX			
PR	09-JUL-1999; 99US-0143257P.		
XX			
PA	(UYN-) UNIV NORTH CAROLINA.		
PA	(ELKI/) ELKINS C.		
XX			
PI	Elkins C;		
XX			
DR	WPI; 2001-138311/14.		
DR	P-PSDB; AAB31712.		
XX			
PT	Novel purified DsrA protein of Haemophilus ducreyi which confers serum		
PT	resistance to the bacteria used to produce vaccines that induce immune		
PT	response against the bacteria in subject at risk of developing chancroid.		
XX			
PS	Claim 1; Page 57-58; 80pp; English.		
XX			
CC	The present sequence encodes a DsrA 30 kDa protein of Haemophilus		
CC	ducreyi. DsrA is an outer membrane protein that confers serum resistance		
CC	to the bacteria. DsrA antisense oligonucleotides are useful for detecting		
CC	a polynucleotide which encodes DsrA in a biological sample. The DsrA		
CC	polypeptide is used to produce vaccine compositions, which are useful for		
CC	inducing a protective immune response in a subject at risk of developing		
CC	Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease		
CC	transmitted by sexual contact. DsrA, its catalytic or immunogenic		
CC	fragments or peptides can be used for screening libraries of compounds in		
CC	a variety of drug screening techniques. The proteins and peptides may be		
CC	used as antigens in immunoassays for detection of Haemophilus ducreyi in		
CC	various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .		
CC	The nucleic acids are useful for the preparation of DsrA proteins		
XX			
SQ	Sequence 1047 BP; 339 A; 163 C; 180 G; 365 T; 0 U; 0 Other;		

Query Match 74.2%; Score 866.8; DB 4; Length 1047;
Best Local Similarity 94.4%; Pred. No. 2.2e-178;
Matches 950; Conservative 0; Mismatches 7; Indels 49; Gaps 3;

```
QY 3 AAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 62
DB |||
QY 42 AATACCGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 101
DB |||
QY 63 CAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 122
DB |||
QY 102 CAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 161
DB |||
QY 123 CCGTAGTGGGATGAGTCTTCTACTATTCAACAATGGCTCAGCAGCGCCAAAGTTTG 182
DB |||
QY 162 CCGTAGTGGGATGAGTCTTCTACTATTCAACAATGGCTCAGCAGCGCCAAAGTTTG 221
DB |||
QY 183 CTGGAGTATCTTTCTGTATAGCTATGAGTATGACTATGTAAGGGTAAATTTGACTTTGGT 242
DB |||
QY 222 CTGGAGTATCTTTCTGTATAGCTATGAGTATGACTATGTAAGGGTAAATTTGACTTTGGT 281
DB |||
QY 243 CTAATGAAGCGGTTTCGATTAATAAGTGCAGGGATTAATAATGAAGCCAAAGATGGA 302
DB |||
QY 282 CTAATGAAGCGGTTTCGATTAATAAGTGCAGGGATTAATAATGAAGCCAAAGATGGA 341
DB |||
QY 303 TTCTCAACAGGCTACTTCTTGAATTAACAGTGCAGGGATTAATAATGAAGCCAAAGATGGA 362
DB |||
QY 342 TTCTCAACAGGCTACTTCTTGAATTAACAGTGCAGGGATTAATAATGAAGCCAAAGATGGA 401
DB |||
QY 363 TGACATATGCTCTGGCGTTTCTCTAG-----CCCTATACCTGTATATCCGATGCTG 416
DB |||
QY 402 TGACATATGCTCTGGCGTTTCTCTAGCTCTATCTCTATACCTGTATATCCGATGCTG 461
DB |||
QY 417 ATCTGTATCACTTGGATTAATCGGAGCAGCTGAAATTTGAATTTGTATAGTTTATTTTA 476
DB |||
QY 462 ATCTGTATCACTTGGATTAATCGGAGCAGCTGAAATTTGAATTTGTATAGTTTATTTTA 521
DB |||
QY 477 ACGATTTAAGACAGATTTTAAATTAAGTCTTGTATGACAGTATTTCCAAATTAAC 536
DB |||
QY 522 ACGATTTAAGACAGATTTTAAATTAAGTCTTGTATGACAGTATTTCCAAATTAAC 581
DB |||
QY 537 AAAATATTGTATATTAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATCTT 596
DB |||
QY 582 AAAATATTGTATATTAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATCTT 641
DB |||
QY 597 ATCGTATGATGGAAC-----AAA 614
DB |||
QY 642 ATCGTATGATGGAACAAATACACATAATCAATAAATAACATAATATCAATAA 701
DB |||
QY 615 ATACACATAATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAG 674
DB |||
QY 702 ATACACATAATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAG 761
DB |||
QY 675 CATTGTCTATGTTAGTGCACCAATGTTGTAGGCAAAAGAGCGTTTCTGCTGCGGTAG 734
DB |||
QY 762 CATTGTCTATGTTAGTGCACCAATGTTGTAGGCAAAAGAGCGTTTCTGCTGCGGTAG 821
DB |||
QY 735 GAGGTTATAGAGATAAAATGTCATTAGCCATTGGTGTGCGCTCAGCGCATTTACTGATCGCT 794
DB |||
QY 822 GAGGTTATAGAGATAAAATGTCATTAGCCATTGGTGTGCGCTCAGCGCATTTACTGATCGCT 881
DB |||
QY 795 TTACCGCTAAAGCGGTTGAGCGTTCAATACCTCAATGCGGCGATGCTTATGCGCTT 854
DB |||
QY 882 TTACCGCTAAAGCGGTTGAGCGTTCAATACCTCAATGCGGCGATGCTTATGCGCTT 941
DB |||
QY 855 CTGTTGGTTATGAATTTCAATCATTTAGTTTAAATCACTAATCGTTTGGTTTATATAAAA 914
DB |||
QY 942 CTGTTGGTTATGAATTTCAATCATTTAGTTTAAATCACTAATCGTTTGGTTTATATAAAA 1001
DB |||
QY 915 AGGCTAAATGTTTCTCTCACATTTAGCC-TTTCTTATTTATCTTT 959
DB |||
QY 1002 AGGCTAAATGTTTCTCTCACATTTAGCC-TTTCTTATTTATCTTT 1047
DB |||
```

RESULT 7

AAF25264

ID AAF25264 standard; DNA; 952 BP.

XX

```
AC AAF25264;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA protein from strain CIPA77.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease; ss.
XX
OS Haemophilus ducreyi.
XX
FH Key Location/Qualifiers
FT CDS 135..929
FT FT /*tag= a
FT FT /product= "DsrA protein"
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (JUNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI; 2001-138311/14.
DR P-PSDB; AAB31707.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Page 53-54; 80pp; English.
XX
CC The present sequence encodes a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc..
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 952 BP; 311 A; 146 C; 172 G; 323 T; 0 U; 0 Other;
```

```
Query Match 71.9%; Score 839.6; DB 4; Length 952;
Best Local Similarity 96.7%; Pred. No. 1.7e-172;
Matches 888; Conservative 0; Mismatches 4; Indels 26; Gaps 2;
```

```
QY 1 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT 60
DB |||
QY 40 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT 94
DB |||
QY 61 TACAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT 120
DB |||
QY 95 TACAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT 154
DB |||
QY 121 TGCCGTAGTGGGATGAGTCTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
DB |||
QY 155 TGCCGTAGTGGGATGAGTCTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 214
DB |||
QY 181 TGCTGGAGTATCTTCTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATG 240
DB |||
QY 215 TGCTGGAGTATCTTCTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATG 274
DB |||
```

```
Qy 241 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAATGAAGCCAAAGAATG 300
Db 275 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAATGAAGCCAAAGAATG 334
Qy 301 GATTCTTAACAGGCTACTTACTTGTAATACAGATTAATGCTTATATCTCTGTCT 360
Db 335 GATTCTTAACAGGCTACTTACTTGTAATACAGATTAATGCTTATATCTCTGTCT 394
Qy 361 CGTGACATATCTCTCGCGTTCTCTAGCCCTATCTACTGTTATATCGATGCTGATCC 420
Db 395 CGTGACATATCTCTAGCGTTCTCTAGCTCTATCTACTGTTATATCGATGCTGATCC 454
Qy 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAAATTTGTATAGTTATTTAAACGA 480
Db 455 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAAATTTGTATAGTTATTTAAACGA 514
Qy 481 TTTAAGACACGATTTTAAATTTAAAGTTCTTGATGCGAGCTATTTCCAAAAATAAACAAA 540
Db 515 TTTAAGACACGATTTTAAATTTAAAGTTCTTGATGCGAGCTATTTCCAAAAATAAACAAA 574
Qy 541 TATTGATCTATAAGTAAATTTTACTAGACTGGTACTTATTAGATGATTTCTTATCG 600
Db 575 TATTGATCTATAAGTAAATTTTACTAGAACTGGTACTTATTAGATGATTTCTTATCG 634
Qy 601 TATGATGGAAC-----AAAAACACATAATATCAATAAGTTGTGTC 639
Db 635 TATGATGGAACAAATACACATAATATCAATAAGTAAATCAATAAGTTGTGTC 694
Qy 640 TAAAGAAATTGCAACTGTTTGTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACAAA 699
Db 695 TAAAGAAATTGCAACTGTTTGTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACAAA 754
Qy 700 TGCTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGGTTATAGATTAATCAATTCGATT 759
Db 755 TGCTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGGTTATAGATTAATCAATTCGATT 814
Qy 760 AGCCATTGGTGTGCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTT 819
Db 815 AGCCATTGGTGTGCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTT 874
Qy 820 CAATACCTTAAATGCGGCGATGCTTATGFGCTCTGTTGGTGTATGAATCTTAATCAAT 879
Db 875 CAATACCTTAAATGCGGCGATGCTTATGFGCTCTGTTGGTGTATGAATCTTAATCAAT 934
Qy 880 ACGTTTAACTCAATATCG 897
Db 935 ACGTTTAACTCAATATCG 952
```

RESULT 8

AAF25265
ID AAF25265 standard; DNA; 899 BP.

XX AC AAF25265;

XX DT 30-APR-2001 (first entry)

XX DE Nucleotide sequence of the DsrA protein from strain CIP542 (Can).

XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX KW chancroid disease; genital ulcer disease; ss.

OS Haemophilus ducreyi.

XX FH Key Location/Qualifiers

XX FT CDS 139..864

XX FT /*tag= a

XX FT /product= "DsrA protein"

XX FT /note= "no termination codon given"

XX FN WO200104138-A1.

XX

PD 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UUNC-) UNIV NORTH CAROLINA.

PA (BLKJ/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

DR P-PSDB; AAB31708.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum

resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.

XX Claim 1; Page 54; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of Haemophilus ducreyi in various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .

XX The nucleic acids are useful for the preparation of DsrA proteins

XX SQ Sequence 899 BP; 291 A; 136 C; 163 G; 309 T; 0 U; 0 Other;

Query Match 71.0%; Score 829.2; DB 4; Length 899;

Best Local Similarity 99.6%; Pred. No. 3e-170;

Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATGTGACATTTTAAATGTAGGTAGTAATAAGAAAGTAAATTCATATT 60

Db 39 ATAAATACGTCATGTGACATTTTAAATGTAGGTAGTAATAAGAAAGTAAATTCATATT 98

Qy 61 TACAATCAAGATTGACAAATTTTAAATGTAGGTAGTAATAAGAAAGTAAATTCATATT 120

Db 99 TACAATCAAGATTGACAAATTTTAAATGTAGGTAGTAATAAGAAAGTAAATTCATATT 158

Qy 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAACAATGGCTCAGCGCGCCAAAGTT 180

Db 159 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAACAATGGCTCAGCGCGCCAAAGTT 218

Qy 181 TGCTGGAGTATCTTTGTATAGCTATGACTATGCTATGCTTAAGGTAATCGACTTG 240

Db 219 TGCTGGAGTATCTTTGTATAGCTATGACTATGCTTAAGGTAATCGACTTG 278

Qy 241 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAATGAAGCCAAAGAATG 300

Db 279 GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAATGAAGCCAAAGAATG 338

Qy 301 GATTCTTAAACAGGCTACTTATCTTGAATTAACAGCATTATATGCTTTATCTCTGTCT 360

Db 339 GATTCTTAAACAGGCTACTTATCTTGAATTAACAGCATTATATGCTTTATCTCTGTCT 398

Qy 361 CGTGACATATGCTCTGGCGTTCTCTAGCCCTATCTGTTATATCGATGCTGATCC 420

Db 399 CGTGACATATGCTCTGGCGTTCTCTAGCCCTATCTGTTATATCGATGCTGATCC 458

Qy 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGTAATTTGTATAGTTATTTAAACGA 480

Db 459 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGTAATTTGTATAGTTATTTAAACGA 518

Qy 481 TTTAAGACACGATTTTAAATTTAAAGTCTTGTATGACGATTTTCCAAAAATAAACAAA 540


```

Db 519 TTTAAGACACAGTTTAAATTAAGTTCTTGATGACAGTATTTCCAAAATAAACAAA 578
Qy 541 TATTGATCTATAAGTAAATTTTACAGAACTGGGTACTTATTAGATGATCTTATCG 600
Db 579 TATTGATCTATAAGTAAATTTTACAGAACTGGGTACTTATTAGATGATCTTATCG 638
Qy 601 TATGATGACAAACAAATACACATAATATCAATAAGTCTCTAAAGAAATTCGAACTGGTTT 660
Db 639 TATGATGACAAACAAATACACATAATATCAATAAGTCTCTAAAGAAATTCGAACTGGTTT 698
Qy 661 AGCCAAACCAATCAGCAATGTCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGCGT 720
Db 699 AGCCAAACCAATCAGCAATGTCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGCGT 758
Qy 721 TTCTGCTGCGGTAGGAGGTTATAGAGATAAACHTGCATTTAGCCATTTGGTGTGGCTCAGG 780
Db 759 TTCTGCTGCGGTAGGAGGTTATAGAGATAAACHTGCATTTAGCCATTTGGTGTGGCTCAGG 818
Qy 781 CATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTACCAATGG 834
Db 819 CATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTTCTATCG 872

```

RESULT 9

AAF25267 standard; DNA; 923 BP.

AC AAF25267;

XX 30-APR-2001 (first entry)

XX Nucleotide sequence of the DsrA protein from strain CH1A.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease; ss.

XX Haemophilus ducreyi.

XX Key Location/Qualifiers

FT CDS 45..836

FT /*tag= a

FT /product= "DsrA protein"

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UNNC-) UNIV NORTH CAROLINA.

XX (BLKJ/) ELKINS C.

XX Elkins C;

XX WPI; 2001-139311/14.

XX P-PSDB; AAB31710.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.

XX Claim 1; Page 56; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins

XX
 SQ Sequence 923 BP; 291 A; 151 C; 174 G; 307 T; 0 U; 0 Other;

Query Match 70.0%; Score 818; DB 4; Length 923;

Best Local Similarity 94.7%; Pred. No. 8.1e-168;

Matches 877; Conservative 0; Mismatches 25; Indels 24; Gaps 2;

Qy 57 TATTTACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTATGAAATTAATATGTT 116

Db 1 TATTTACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTATGAAATTAATATGTT 60

Qy 117 TAGTTGCCGTAGTGGGATTAGCTTTGTTTCTACTATTACAACAATGGCTCAGCAGCGCCAA 176

Db 61 TAGTTGCCGTAGTGGGATTAGCTTTGTTTCTACTATTACAACAATGGCTCAGCAGCGCCAA 120

Qy 177 AGTTTGGTGGAGTATCTCTTTGTATAGCTATGACTATGACTATGTAAGGGTAATGGA 236

Db 121 AGTTTGGTGGAGTATCTCTTTGGATAGCTATGACTATGTAAGGGTAATGGA 180

Qy 237 CTTTGTCTAATGAAGCGGTTTCGATATTAAAGTGCAGGATTAAATGAAGCCAAAG 296

Db 181 CTTTGTCTAATGAAGCGGTTTCGATATTAAAGCGGATTAAATGAAGCCAAAG 240

Qy 297 AATGGATTCTAAACAGGCTACTTATCTTGAATTACAGCAATTATATGCCCTTATCTCCTG 356

Db 241 AATGGATTCTAGACAGGCTACTTATCTTGAATTACAGCAATTATATGCCCTTATCTCCTG 300

Qy 357 TTCTCTGATGATATGCTCTCGGCTTTCTCTAGCCCTATCTATGTTATATCCGATGCTG 416

Db 301 TTCTCTGATGATATGCTCTCTGCGAG---AACCTAACACGTGACTGTTATATCCGATGCTG 357

Qy 417 ATCTGATCAACTTGGAAATTAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTA 476

Db 358 ATCTGATCAACTTGGAAATTAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTA 417

Qy 477 ACGATTTAAGACACAGATTTTAAATTAATAAGTTCTTGATGCACGTATTTCCAAAATAAAC 536

Db 418 ACGATTTAAGACACAGGTTTAAATTAATAAGTTCTTGATGCACGTATTTCCAAAATAAAC 477

Qy 537 AAAATATTGATCTATAAGTAAATTTTACTAGAACCTGGGTACTTATTTAGATGATCTTT 596

Db 478 AAAATATTGATCTATAAGTAAATTTTACTAAAACTGGGTACTTATTTAGATGATCTTT 537

Qy 597 ATCGTATGATGGAAC-----AAAATACACATAATCAATAAGT 635

Db 538 ATCGTATGATGGAACAAATAACATAATATCAATAATAACATAATCAATAAGT 597

Qy 636 TGTCTAAAGAAATTCGAAACTGGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGAAC 695

Db 598 TGTCTAAAGAAATTCGAAACTGGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGAAC 657

Qy 696 CAAATCGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAAACTG 755

Db 658 CAAATCGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAAACTG 717

Qy 756 CATTAGCCATTGTTGTCGGCTCAGGCATTTACTGATCGCTTTACCGCTTAAGCGGTTGAG 815

Db 718 CATTAGCCATTGTTGTCGGCTCAGGCATTTACTGATCGCTTTACCGCTTAAGCGGTTGAG 777

Qy 816 CGTTCAATACCTCAATGCGGCATGCTTTATCGGCTCTCTGTTGGTTATGAATTTCTAAT 875

Db 778 CGTTCAATACCTCAATGCGGCATGCTTTATCGGCTCTCTGTTGGTTATGAATTTCTAAT 837

Qy 876 CATTAGCTTTAATCAGTAAATCGTTTTGGTTTATATAAAGGCTAAATGTTTCTCCTCAC 935

Db 838 CATTAGCTTTAATCAGTAAATCGTTTTGGTTTATATAAAGGCTAAATGTTTCTCCTCAC 897

```
Qy 936 ATTAGCCTTCTTATTATCTTTGT 961
Db 898 ATTAGCCTTCTTATTATCTTTGT 923

RESULT 10
AAC79587/c
ID AAC79587 standard; DNA; 4426 BP.
XX
AC AAC79587;
XX
DT 08-FEB-2001 (first entry)
XX
DE Virulence gene #7.
XX
KW Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;
KW bronchopneumonia; rhinitis; wound infection; ss.
XX
OS Pasteurella multocida.
XX
PN WO200061724-A2.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US009218.
XX
PR 09-APR-1999; 99US-0128689P.
PR 10-SEP-1999; 99US-0153453P.
XX
PA (PHAA ) PHARMACIA & UPJOHN INC.
XX
PI Lowery DE, Fuller TE, Kennedy MJ;
XX
DR WPI; 2000-647422/62.
DR P-PSDB; AAB44527.
XX
PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
PT genes, useful as a live attenuated vaccine against bacterial infections.
XX
PS Claim 7; Page 83-85; 32pp; English.
XX
CC The family Pasteurellaceae encompasses several pathogens that infect a
CC wide variety of animals. The present invention relates to virulence genes
CC from Pasteurellaceae. The present sequence is one such virulence gene.
CC The present sequence may be mutated in order to produce an inactive gene.
CC The inactive virulence gene may in turn be used to produce a vaccine,
CC which is useful for treating bacterial infections such as septicemias,
CC bronchopneumonias, rhinitis and wound infections
XX
SQ Sequence 4426 BP; 1343 A; 882 C; 827 G; 1374 T; 0 U; 0 Other;

Query Match 5.8%; Score 67.2; DB 3; Length 4426;
Best Local Similarity 70.3%; Pred. No. 7.8e-05;
Matches 90; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 1038 ACTTTCACCTTTTTCACCATATTATCCGCCACTTCTAAAACGGTAATATTAAAGTTGGTTT 1097
Db 2512 ACTTTGGCTTGTTAATCATGTTGTCTATCTTCTAGATCGTGATTTGTAGCCCATTA 2453
Qy 1098 AGCCTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTTCTTAAATTAAGCCGTTA 1157
Db 2452 ATTTACATACCGTTCCTTCTCTGGAATTTCTTCTTAAATGCTCTTAAATTAAGCCGTTG 2393
Qy 1158 AAGGTGCG 1165
Db 2392 AAGGTACG 2385

RESULT 11
ABQ83464/c
ID ABQ83464 standard; DNA; 4426 BP.
XX
AC ABQ83464;
```

```
XX
DT 07-AUG-2003 (revised)
DT 24-JAN-2003 (first entry)
XX
DE Pasteurella multocida exbB gene SEQ ID NO:13.
XX
KW Antibacterial; vaccine; gram negative bacterial virulence gene;
KW identification; virulence; Pasteurellaceae; gene; ds.
XX
OS Pasteurella multocida.
XX
PN WO200275507-A2.
XX
PD 26-SEP-2002.
XX
PF 17-JAN-2002; 2002WO-US001971.
XX
PR 15-MAR-2001; 2001US-00809665.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Lowery DE, Fuller TE, Kennedy MJ;
XX
DR WPI; 2002-740868/80.
DR P-PSDB; ABP54479.
XX
PT New mutant gram-negative bacteria, useful as vaccines and for identifying
PT new anti-bacterial agents that target virulence genes and their products.
XX
PS Claim 5; Page 87-90; 350pp; English.
XX
CC The present invention describes a gram-negative bacteria comprising a
CC mutation in a gene, where the mutation results in decreased activity of a
CC gene product encoded by the mutated gene. Also described is a method for
CC producing a gram-negative bacteria mutant or an attenuated
CC Pasteurellaceae bacteria. The mutated genes have antibacterial activity
CC and can be used in vaccines. The gram-negative bacteria or the attenuated
CC Pasteurellaceae bacteria can be used as vaccines in the fields of human
CC medicine or veterinary medicine, and for identifying new antibacterial
CC agents that target the virulence genes and their products. ABQ83458 to
CC ABQ8578 and ABP54473 to ABP54551 represents sequences used in the
CC exemplification of the present invention. (Updated on 07-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 4426 BP; 1342 A; 882 C; 827 G; 1374 T; 0 U; 1 Other;

Query Match 5.8%; Score 67.2; DB 6; Length 4426;
Best Local Similarity 70.3%; Pred. No. 7.8e-05;
Matches 90; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 1038 ACTTTCACCTTTTTCACCATATTATCCGCCACTTCTAAAACGGTAATATTAAAGTTGGTTT 1097
Db 2512 ACTTTGGCTTGTTAATCATGTTGTCTATCTTCTAGATCGTGATTTGTAGCCCATTA 2453
Qy 1098 AGCCTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTTCTTAAATTAAGCCGTTA 1157
Db 2452 ATTTACATACCGTTCCTTCTCTGGAATTTCTTCTTAAATGCTCTTAAATTAAGCCGTTG 2393
Qy 1158 AAGGTGCG 1165
Db 2392 AAGGTACG 2385

RESULT 12
AAZ54498
ID AAZ54498 standard; DNA; 1095 BP.
XX
AC AAZ54498;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 961 partial DNA sequence SEQ ID NO:2943.
XX
```


Search completed: May 15, 2004, 03:00:25
Job time : 569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 15, 2004, 01:33:10 ; Search time 4881 Seconds
(without alignments)
10371.781 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 1168
Sequence: 1 atcaatagctcattgacatt.....aagcgttaaaggtagggac 1168

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 10%
Maximum Match 100%
Listing first 45 summaries

Database :	1: gb.ba.*	2: gb.htg.*	3: gb.in.*	4: gb.om.*	5: gb.ov.*	6: gb.pat.*	7: gb.ph.*	8: gb.pl.*	9: gb.pr.*	10: gb.ro.*	11: gb.sts.*	12: gb.sy.*	13: gb.un.*	14: gb.vi.*	15: em.ba.*	16: em.fun.*	17: em.hum.*	18: em.in.*	19: em.mu.*	20: em.om.*	21: em.or.*	22: em.ov.*	23: em.pat.*	24: em.ph.*	25: em.pl.*	26: em.ro.*	27: em.sts.*	28: em.un.*	29: em.vi.*	30: em.htg.hum.*	31: em.htg.inv.*	32: em.htg.other.*	33: em.htg.mus.*	34: em.htg.pln.*	35: em.htg.rod.*	36: em.htg.mam.*	37: em.htg.vrt.*	38: em.sy.*	39: em.htgo.hum.*	40: em.htgo.mus.*	41: em.htgo.other.*
	1: gb.ba.*	2: gb.htg.*	3: gb.in.*	4: gb.om.*	5: gb.ov.*	6: gb.pat.*	7: gb.ph.*	8: gb.pl.*	9: gb.pr.*	10: gb.ro.*	11: gb.sts.*	12: gb.sy.*	13: gb.un.*	14: gb.vi.*	15: em.ba.*	16: em.fun.*	17: em.hum.*	18: em.in.*	19: em.mu.*	20: em.om.*	21: em.or.*	22: em.ov.*	23: em.pat.*	24: em.ph.*	25: em.pl.*	26: em.ro.*	27: em.sts.*	28: em.un.*	29: em.vi.*	30: em.htg.hum.*	31: em.htg.inv.*	32: em.htg.other.*	33: em.htg.mus.*	34: em.htg.pln.*	35: em.htg.rod.*	36: em.htg.mam.*	37: em.htg.vrt.*	38: em.sy.*	39: em.htgo.hum.*	40: em.htgo.mus.*	41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1168	100.0	301191	1	AE017152	AE017152 Haemophil
2	1000	85.6	1000	1	AF187001	AF187001 Haemophil
3	874	74.8	874	1	AF187005	AF187005 Haemophil
4	874	74.8	874	1	AF187007	AF187007 Haemophil
5	829.2	71.0	847	1	AF187004	AF187004 Haemophil
6	816.6	69.9	890	1	AF187002	AF187002 Haemophil
7	816.6	69.9	890	1	AF187003	AF187003 Haemophil
8	814	69.7	916	1	AF187009	AF187009 Haemophil
9	792.8	67.9	923	1	AF187008	AF187008 Haemophil
10	780.6	66.8	894	1	AF187006	AF187006 Haemophil
11	84.6	7.2	1967	1	AB064943	AB064943 Actinobac
12	84.6	7.2	2579	1	AF316502	AF316502 Actinobac
13	68.6	5.9	2216	1	YPYOPA	X13883 Yersinia ps
14	68.4	5.9	2216	1	YPYOPANF	X13880 Yersinia pe
15	68.4	5.9	70305	1	YPCD1	AL117189 Yersinia
16	68.4	5.9	70504	1	AF053946	AF053946 Yersinia
17	68.4	5.9	70559	1	AF074612	AF074612 Yersinia
18	68	5.8	66591	1	AY150843	AY150843 Yersinia
19	67.2	5.8	4426	6	BD252123	BD252123 Anti-Bact
20	67.2	5.8	4426	6	AX553679	AX553679 Sequence
21	67.2	5.8	11632	1	AE006158	AE006158 Pasteurel
22	66.4	5.7	2186	1	YEQOPA80	X13881 Yersinia en
23	66.4	5.7	67720	1	AF336309	AF336309 Yersinia
24	63.8	5.5	2551	1	YEQOPA64	X13882 Y. enterocol
25	61.8	5.3	3167	1	AF399847	AF399847 Escherich
26	60.2	5.2	11279	7	AF151091	AF151091 Prophage
27	60	5.1	3795	7	AF151675	AF151675 Bacteriop
28	60	5.1	9689	7	AF151674	AF151674 Bacteriop
29	59.8	5.1	1760	1	AF452479	AF452479 Neisseria
30	59.8	5.1	1761	1	AF452481	AF452481 Neisseria
31	59.8	5.1	1765	1	AF452477	AF452477 Neisseria
32	59.8	5.1	1767	1	AF452472	AF452472 Neisseria
33	59.8	5.1	1778	1	AF452486	AF452486 Neisseria
34	59.8	5.1	1779	1	AF452466	AF452466 Neisseria
35	59.8	5.1	1782	1	AF452467	AF452467 Neisseria
36	59.8	5.1	1789	1	AF452478	AF452478 Neisseria
37	59.8	5.1	1791	1	AF452485	AF452485 Neisseria
38	59.8	5.1	1794	1	AF452475	AF452475 Neisseria
39	59.8	5.1	9977	1	AE002548	AE002548 Neisseria
40	59.8	5.1	172325	6	AX044035	AX044035 Sequence
41	59.8	5.1	349980	6	AX044034	AX044034 Sequence
42	59.6	5.1	10903	1	AE006108	AE006108 Pasteurel
43	59.2	5.1	972	6	AX698453	AX698453 Sequence
44	59.2	5.1	1870	1	AF452484	AF452484 Neisseria
45	59.2	5.1	1874	1	AF452476	AF452476 Neisseria

ALIGNMENTS

RESULT 1	AE017152	301191 bp	DNA	linear	BCT 15-AUG-2003
LOCUS	Haemophilus ducreyi strain 3500HP				section 2 of 6 of the complete genome.
DEFINITION	AE017152				
ACCESSION	AE017152				
VERSION	AE017152.1	GI:33147833			
KEYWORDS	Haemophilus ducreyi	3500HP			
SOURCE	Haemophilus ducreyi	3500HP			
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.				
REFERENCE	1 (bases 1 to 301191)				
AUTHORS	Munson, R.S. Jr., Ray, W.C., Mahairas, G., Sabo, P., Mungur, R., Johnson, L., Nguyen, D., Wang, J., Forst, C. and Hood, L.				

Pred. No. is the number of results predicted by chance to have a

```

TITLE       The Complete Genome Sequence of Haemophilus ducreyi
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 301191)
AUTHORS     Munson,R.S. Jr., Ray,W.C., Mahairas,G., Sabo,P., Mungur,R.,
            Johnson,L., Nguyen,D., Wang,J., Forst,C. and Hood,L.
TITLE       Direct Submission
JOURNAL     Submitted (04-JUN-2003) Pediatrics, Columbus Children's Research
            Institute and The Ohio State University, 700 Children's Drive,
            Columbus, OH 43205, USA
FEATURES    Location/Qualifiers
             source
               1..301191
               /organism="Haemophilus ducreyi 35000HP"
               /mol_type="genomic DNA"
               /strain="35000HP"
               /db_xref="taxon:233412"
               115..1143
               /gene="apbE"
               /locus_tag="HD0386"
               115..1143
               /gene="apbE"
               /locus_tag="HD0386"
               /codon_start=1
               /transl_table=11
               /product="thiamine biosynthesis lipoprotein"
               /protein_id="AAP95355.1"
               /db_xref="GI:33147834"
               /translations="MLVCVASLSFILTACNKSAEQITLQKTMGTTVTYVKYIDGKIA
               NLTAEOVKQLDDVLKVNDEMSTYQLDSQISRFNKLRENTVPFELSPDPATVAEA
               IRLTKITGALDVTGVLNMGFGPDRUDKQPSAAQINORAAFVGIDIKLANIEHK
               PTNKKYTVLDDSSIAKGFGVDKIAQDLEMGLGLENLYVEVGELRGKGNKLGNNW
               QIAIEKNFPGQGTQVITVPLHNLGMAISGNYRYFEQGNRLSHIIDPKALKPVNH
               KLASITVLATSMTADGLTGLFVLGAEKALEVAEREKLAIFLIKNKGDFGTQMSSQ
               FKXLDQK"
               1155..1403
               /locus_tag="HD0387"
               1155..1403
               /locus_tag="HD0387"
               /codon_start=1
               /transl_table=11
               /product="conserved hypothetical protein"
               /protein_id="AAP95356.1"
               /db_xref="GI:33147835"
               /translations="METLLTTFGFFILVIFAMSIGFLIKNGISCGGITALGMKMN
               CDCEPCDNLTKIMEGKADPAKIAQPKKVASPEVK"
               complement(1459..3678)
               /gene="tdhA"
               /locus_tag="HD0388"
               complement(1459..3678)
               /gene="tdhA"
               /locus_tag="HD0388"
               /codon_start=1
               /transl_table=11
               /product="ton-dependent heme receptor A"
               /protein_id="AAP95357.1"
               /db_xref="GI:33147836"
               /translations="MKMKQCATLTFIFLGHVYTIADNPKNISLSVITVPGHHEROP
               DRSLITONEIDQSDNVADLVNTVPGVSMAGGPRPSGQTLNIRMGDGTEDIRVOVDG
               ATKNFEXQOQSLEPIEELLRRYSIDKGNHYPOYGNGGAGTIKLETKNAKDFLENQ
               LUGLLKGYTNNNQRTFSGALPMQNDQKNIDALVYATVRRAGTIKRAKDPIKYS
               NQANFLAKVNWMLTSQLAFKSVGNHNGWEPFAAKRDLPLGPTAEITKYGLDIA
               WKGLAREQDQRSYSIQYQFLPENNPWINTVAQLSHSSTYQDTSQASKYTLASL
               GNSSTRYTDLTFDNNYSLFNVAKTSHTLLVGLQWKHKRQTLIPDPSKLOAEVNH
               GYQPSYMPSGOYTHAFYADQDKIKHNLTVSICARVDYVKNIGKPNIAITVNDPTAG
               HDVSSKHPGKSSVLGNLYKLTPLYNLFSNISNTRAPVIDEQYETOYAKATLSPTAS
               SLQDKRITQLRVGKQIHFDHILSNNDQLSFNSFTFYFGKDEIFKTRGVRCPESAQ
               NNNNEVKIKIGNRYNLPGYQIKGFEALYNDSTYFNLSYSHTIKRLASPRNPWL
               ASTSWIAEIPPRKAVVTLGSHIPDTNLTLGWKSEFVRQDRSPSTDQDKAGHALPKS
               SGVALHGIFATWQKIKHLRIQFTVDNLLNRSYRPSYLSLAAGTGRNLIKLSISKQF"
               3881..5035
               /gene="trmU"
               /locus_tag="HD0389"
               3881..5035
               /gene="trmU"

```

```

/locus_tag="HD0389"
/codon_start=1
/transl_table=11
/product="tRNA
(S-methylaminomethyl-2-thiouridylate)-methyltransferase"
/protein_id="AAP95358.1"
/db_xref="GI:33147837"
/translations="MALNSITYQKCFHTSAQQYADNKKKVIIGMSGGVSSVSFAFIL
QQQGYOVGLFMKNWEEDDDTDYCTAAADLADQNVADRLGILKHKINFAAEYWDNVF
EHELASYQAGRTPNPILCNKEIKFALEVAEADLADYIATGCHYVRSADNENAOQL
RGLDANKQDSYFLYTLNSQVAKSLFFVEIEKPMWRQIATDGLVTAKKDSGTGIC
FIGERKPEFLARFLPAQPGDIDSTVEGIEVGRHDGLMYTYTLGQRKGLIGIGVKGSLSD
PFYVVEKDLVNNYLIVAQQHDNGALLSTGLIANQLHWDVDRKTIQDTRCTVTKTYRQT
DIACLVEPIDAENIRVIFDPEQIAVTPGQSAVFQSGVCLGGGVIEAQLK"
complement(5103..6257)
/gene="hcat"
/locus_tag="HD0391"
complement(5103..6257)
/gene="hcat"
/locus_tag="HD0391"
/codon_start=1
/transl_table=11
/product="probable 3-phenylpropionic acid transporter"
/protein_id="AAP95359.1"
/db_xref="GI:33147838"
/translations="MIKPTPQWTAQNFPGFCAYGVLMPFLPVWLRRHHGYTTEMIGL
LASFGYLFAPAGGMLSSQAWAKOANQLIPTLRRLTLNIMTIILLAYTVDSVWLLFPAL
MLPHINACTMPTISDSIASLWQKQIGLDYGKARLPGSIAFVSGSLTGYLANWGEQV
IIMWAGFFLLSGQQLTTPQHFVNNQAQCHUTYQWLKQPTTLRMMIALSLMLG
GSHAAYTSTIHWSTAGISTOTSSFLMGTAVCSEIIFFLFANQLFKSWRIQHLMLIG
SIAAIIRWGSLASTHTHVALFCSTQTHALTFGVAHVAMIRYISTQASEKVTYKLGQLYF
SIANCCGFMALFTFESIGTYIQDLNFAPFALVAPAILVPKQFPVNLK"
6400..7302
/gene="rdgC"
/locus_tag="HD0392"
6400..7302
/gene="rdgC"
/locus_tag="HD0392"
/codon_start=1
/transl_table=11
/product="recombination associated protein"
/protein_id="AAP95360.1"
/db_xref="GI:33147839"
/translations="MFWKFNMIYRLTNALQLASNELEPLLQOHRFGCSQSEMSKFG
WTAPLAKTEQLHFCHTKQFLLI SHKEDKLPAVNIKAETERRIALLSEKNRKLKQTE
KQAIKDDVTAMLPRAFSKHQFTALWLDLDAQLVYVDAASSKRAEASIALLRKTLGSI
FVVPFSVQLPFAEVTNWLAKGHPNMTLLEAEALSKFTESVIRCKKQELSDSEIC
YHLOAGKLVTKLAIEWENHPSCVLNNDATILRLKPAGEIIEKNDIDILKEDIQAQRLAD
FLIMTELRFLTQKLIIEFGGIKE"
complement(7413..9194)
/locus_tag="HD0393"
complement(7413..9194)
/locus_tag="HD0393"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP95361.1"
/db_xref="GI:33147840"
/translations="MIRLQPNQLKDKLNFTIEQVLNKAADAAGSKMDANANITQKNIAI
MENEMKDFVQINRAKVNKPIAEFLDAELASELYRQIEHHEIYVHDETSLKPYCVSV
TWYFRLRYGLTKFGSGSKNPKILASFCGSFNVFVFAVSQFAGAVATVEFLTYFDYFA
RKDFGDYLLTBHKKQECNLHQVVYVSINQPAAGVQSOFVFNISLYDKYIPEAMFSOF
VFDPFRPNQTTAKQLPFMKWFOERTKAILTFPVI TAMLTHEQCKQDNQFADOM
AEBSQGNSEFFIYQSDNPDSLASCCLRNIEADHTFSYSLGAGSVATGSINVTIWNK
RLVQDRDLAKEVEKHKHYAYRKLAMEDYLAAGMLPAYDAGFTILKQDQFVITTINGM
AAEAASQGLTVGYNQYIYDFVSRNLNVI FVANQTASQQYGVKNEFTFVPAENLGVKXA
KWDQSGLYFPRDCYNSYFYLVQDQINALDKFLHKGQLIEWLDGSGALHNLDESIL
SKYGYRSLDLAAKTCNFCVNVKITICNECHYIDKRTLHHCCKGCSANIDYGTTRI
GYLKYRSAPSSARQBSLRFYHQAA"
complement(9194..9667)
/locus_tag="HD0394"
complement(9194..9667)
/locus_tag="HD0394"

```

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP95362.1"
/db_xref="GI:33147841"
/translation="MNSIRFCSQIVQVEVPRETSIAFLITGTCPLGCKGCHSDSWKA
GSOIUTSVLAELAKYDLRLISCVPMGEWHLOKLLSLLKLVKQAGLHTCLYTGLE
QSLPADLAELIYLTATGRWIAERGLTSLTNNQHFIDLTGLENLHFIKEFER"
10039..10620
/gene="rnfA"
/locus_tag="HD0396"
/locus="rnfA"
/locus_tag="HD0396"
/codon_start=1
/transl_table=11
/product="probable Na+-translocating NADH-quinone
oxidoreductase"
/protein_id="AAP95363.1"
/db_xref="GI:33147842"
/translation="MWDYILLYVSTALINFLVFLVFLGLCPFMGVSKVETAIQMGA
TFVFLTASLVSAYLVERFLIPIDQFLRLTFLVILVIAIVQTEMIVHKTSPTLRL
LGYLPLITNCVAVGLVALLNLSNLLIESVLVYGFAGAVGSLVFLVLAALRERLAA
ADVRPFQASIALITAGLMSLAFMGFTGLVKL"
10620..11243
/gene="rnfB"
/locus_tag="HD0397"
/locus="rnfB"
/locus_tag="HD0397"
/locus="rnfB"
/locus_tag="HD0397"
/codon_start=1
/transl_table=11
/product="putative ferredoxin II, iron sulfur protein"
/protein_id="AAP95364.1"
/db_xref="GI:33147843"
/translation="MLDLSIIAYLLIAILIFGALLGYFSVKLVKVEADPIVDI
AIIPOSGCGGPGKPAEATANGDOITKVPQGPLVVKIAELMGLVDVPSMDNTA
VPEVKVALLHEDWICIGCTKCIQACPVDAILGYNKAMHTVVDLCTGCELCVAPCPNC
IEMIKVQKPRSNWQFNPDLLIPIVNTTELQKQIMVGSAGEIQND"
11236..11344
/gene="rnfC"
Query Match 100.0%; Score 1168; DB 1; Length 301191;
Best Local Similarity 100.0%; Pred. No. 8.1e-206; Gaps 0;
Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATAATACGTCATGTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 60
DB 298759 AATAATACGTCATGTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 298818
QY 61 TACAATCAAGATTGACAAATTTATTTACTTAATAGAGGTGATTATGAAATTAATGTTTAGT 120
DB 298819 TACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 298878
QY 121 TCGCGTAGTGGGATTAGCTGTCTACTATTACAAGTGGCTCAGCGCGCAAGTT 180
DB 298879 TCGCGTAGTGGGATTAGCTGTCTACTATTACAAGTGGCTCAGCGCGCAAGTT 298938
QY 181 TCGTGGAGTATCTTCTTTGATAGCTATGATGATGACTATGTTAAAGGTAAGTGGACTTG 240
DB 298939 TCGTGGAGTATCTTCTTTGATAGCTATGATGATGACTATGTTAAAGGTAAGTGGACTTG 298998
QY 241 GTCTAATGAAGCGGTTTCGATATAAAGTGCCAGGGATTAAATGAAGCCAAAGATG 300
DB 298999 GTCTAATGAAGCGGTTTCGATATAAAGTGCCAGGGATTAAATGAAGCCAAAGATG 299058
QY 301 GATTCTAAACAGGCTACTTATCTTGAATTTACAGCATTAATGCTTATCTGTTCT 360
DB 299059 GATTCTAAACAGGCTACTTATCTTGAATTTACAGCATTAATGCTTATCTGTTCT 299118
QY 361 CGTGACATATGCTCTGGCGGTTCTCTAGCCCTTACTGTTATATCCGATGCTGATCC 420
DB 299119 CGTGACATATGCTCTGGCGGTTCTCTAGCCCTTACTGTTATATCCGATGCTGATCC 299178
```

```
QY 421 TGATCAACTTGGAAATAAATCGCAGCAGCTGAAATTTGAATTTGTATAGTATTTTAAACGA 480
DB 299179 TGATCAACTTGGAAATAAATCGCAGCAGCTGAAATTTGAATTTGTATAGTATTTTAAACGA 299238
QY 481 TTTAAGACACGATTTTAAATTTAAAGTCTTCTGATGCACGTATTTCCAAAATAAACAATAA 540
DB 299239 TTTAAGACACGATTTTAAATTTAAAGTCTTCTGATGCACGTATTTCCAAAATAAACAATAA 299298
QY 541 TATTGATCTAATAAGTAATATTTACTAGAACTGGGTACTTATTTAGATGATCTTATCG 600
DB 299299 TATTGATCTAATAAGTAATATTTACTAGAACTGGGTACTTATTTAGATGATCTTATCG 299358
QY 601 TATGATGGAACAAAATACACATAATCAATCAATAGTTGTCTAAAGAATTCGAAACTGGTTT 660
DB 299359 TATGATGGAACAAAATACACATAATCAATCAATAGTTGTCTAAAGAATTCGAAACTGGTTT 299418
QY 661 AGCCAACCAATCAGCAATCTCTATGTTAGTCAACCAAAATGGTGTAGGCCAAAACGAGCGT 720
DB 299419 AGCCAACCAATCAGCAATCTCTATGTTAGTCAACCAAAATGGTGTAGGCCAAAACGAGCGT 299478
QY 721 TTCCTGTCGGGTAGGAGGTATAGAGATAAAACTGCAATAGCCATTTGGTGTGGCTCAGC 780
DB 299479 TTCCTGTCGGGTAGGAGGTATAGAGATAAAACTGCAATAGCCATTTGGTGTGGCTCAGC 299538
QY 781 CATTACTGATCGCTTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAAATCGCGGCAT 840
DB 299539 CATTACTGATCGCTTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAAATCGCGGCAT 299598
QY 841 GTCCTATGCTGCTTCTGTTGGTTATGAATTTCAATCATTACGTTTAAATCACTAATCGTTT 900
DB 299599 GTCCTATGCTGCTTCTGTTGGTTATGAATTTCAATCATTACGTTTAAATCACTAATCGTTT 299658
QY 901 TGGTTATAATAAAAGCGCTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 960
DB 299659 TGGTTATAATAAAAGCGCTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 299718
QY 961 TTATAGCTTTTCTGTTTATAAAACCGTTTCTTAGCCACATTTTATTAATAAGCTTTTAAAG 1020
DB 299719 TTATAGCTTTTCTGTTTATAAAACCGTTTCTTAGCCACATTTTATTAATAAGCTTTTAAAG 299778
QY 1021 CCTATTCAATCAGTCTTCTACTTTTACCTTTTCCACCATTTATTCGCCACTTCTTAAACGG 1080
DB 299779 CCTATTCAATCAGTCTTCTACTTTTACCTTTTCCACCATTTATTCGCCACTTCTTAAACGG 299838
QY 1081 TAATAATTAAGTTGGTTTACCTTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTT 1140
DB 299839 TAATAATTAAGTTGGTTTACCTTAAATGGGTACCTTCTATCGGAATTTTCTTAAATGTT 299898
QY 1141 CTAAATTAAGCGGTTAAAGGTGCGGAC 1168
DB 299899 CTAAATTAAGCGGTTAAAGGTGCGGAC 299926
```

RESULT 2

```
AF187001
LOCUS AF187001
DEFINITION Haemophilus ducreyi strain 35000 serum resistance protein Dsra
(dsra) gene, complete cds.
ACCESSION AF187001
VERSION AF187001.1 GI:7188572
KEYWORDS
SOURCE
ORGANISM Haemophilus ducreyi
Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
1 (bases 1 to 1000)
REFERENCE
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein Dsra
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
```

REFERENCE 2 (bases 1 to 1000)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
Location/Qualifiers
1..1000
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="35000"
/db_xref="taxon:730"
/note="serum resistant strain"
101..874
/gene="dsrA"
101..874
/gene="dsrA"
/note="outer membrane protein required for serum resistance"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37807.1"
/db_xref="GI:7188573"
/translation="MKIKCLVAVVGLACSTITWAOQPPKPGVSSLYSEYDYGKKG
WTNSGEGFDIKVPGIKMKPKEMISKQATYLELQHYMPYTVLVTPYAPGVSPILLY
PMSDPDOLGNROQLKLNLYSYFNDLRHDFKLVDARISKVNKONIDTISKYLLLELGT
YLDPSYRMEQNTNHNKLSKELDTGLANOSALSMVQPNVGKTSVSAVGVYRDKT
ALATGVSRTDRFTAKGVAFNYNGSMYSVASVGYEF"

ORIGIN
Query Match 85.6%; Score 1000; DB 1; Length 1000;
Best Local Similarity 100.0%; Pred. No. 2.3e-174;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAAATAAGAAAGTAATTTCTATTT 60
Db |||||||
Qy 61 TACAATCAAGATTGACAAATTTTAAATGTAAGGTAGAAATAAGAAAGTAATTTCTATTT 120
Db |||||||

Qy 121 TCGCGTAGTGGGATTAGCTTCTTCTACTATTACAAATGGCTCAGACGCGCCAAAGTT 180
Db |||||||

Qy 181 TGCTGGAGTATCTTCTTTGTATAGCTATGATGTGACTATGGTAAGGGTAAATGGACTTG 240
Db |||||||

Qy 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAATAATGAAGCCAAAGATG 300
Db |||||||

Qy 301 GATTTCTAAACAGGCTACTTATCTTTGAATTACAGCAATTATATGCTTATATCTCTGTCT 360
Db |||||||

Qy 361 CGTGACATATGCTCTCGCGTTTCTCTAGCCCTATCTGTATATCCGATGTCGTATCC 420
Db |||||||

Qy 421 TGATCACTTGGATTAATCGCAGCTCAATTCGAATTCGAATTCGATGTTATTTTACGA 480
Db |||||||

Qy 481 TTTAAGACACGATTTTAAATTTAAAGTTCTTGTATGTCAGCTATTTCCAAAAATAACAAA 540
Db |||||||

Qy 541 TATTGATCTATAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTCCTATCG 600
Db |||||||

Db 541 TATTGATCTATAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTCCTATCG 600
Qy |||||||

Db 601 TATGATGGACAAAATACACATATATCAATAAGTTTCTCTAAAGAAATGCAAACTGGTTT 660
Db |||||||

Qy 601 TATGATGGACAAAATACACATATATCAATAAGTTTCTCTAAAGAAATGCAAACTGGTTT 660
Db |||||||

Qy 661 AGCAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
Db |||||||

Qy 721 TTCTGTCGGTAGGAGGTTATAGAGATAAATGCAATAGGCATTTGGTTCGGCTCACG 780
Db |||||||

Qy 781 CATTACTGATCGCTTTACCGCTAAAGCGGCTGAGCGTTCAATACCTACCAATGGCGGCAT 840
Db |||||||

Qy 841 GTCTTATGGTCTTCTGTTTATGAATTTCTAATCATTACGTTTAAATCACTAATCGTTT 900
Db |||||||

Qy 901 TGGTTATAATAAAGCTAAATGTTTCTCTCCTCACATTTAGCCCTTCTTATTTATCTTTG 960
Db |||||||

Qy 961 TTATAGCTTTTGTCTGTTTATAAAACCGTTTTTTTAGCCACTT 1000
Db |||||||

RESULT 3
AF187005 874 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain 406 serum resistance protein DsrA (dsrA)
DEFINITION gene, complete cds.
ACCESSION AF187005
VERSION AF187005.1 GI:7188580
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 874)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 874)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
Location/Qualifiers
1..874
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="406"
/db_xref="taxon:730"
/note="serum resistant strain"
101..874
/gene="dsrA"
101..874
/gene="dsrA"
/note="outer membrane protein required for serum resistance"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37811.1"

```
/db_xref="GI:7188581"
/translation="MKIKCLVAVVGLACSTTTMAQOPKFAVSSLYSEYDYGKCK
WTMSNEGDFIKVPKIKMKPKWISKOATYLELOHYMPYPLVTVYAGVSPFLLY
PMSDDQGLNROQLKUNLYSYFNDLRHDFKLKVLDIRISKNQKNDITISKYLLLEGT
YLDSSYRMQONTHNINKLSKELOQTGLANQSLMLVOPNGVKGTSVSAAGVGRDXT
ALAIGVSRITDRFTAKAGVAFNTYNGMSYGASVGYEF"

ORIGIN
Query Match      74.8%; Score 874; DB 1; Length 874;
Best Local Similarity 100.0%; Pred. No. 3.8e-151;
Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60
Db 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60

Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120
Db 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120

Qy 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
Db 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180

Qy 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240
Db 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240

Qy 241 GTCTATGAGCGGTTTCGATATTAAAGTCCAGGATTAATTAAGTAAAGTAAAGTAAAGT 300
Db 241 GTCTATGAGTCTTCTGTTGTTATGAATTCCTAA 874

ORIGIN
Query Match      74.8%; Score 874; DB 1; Length 874;
Best Local Similarity 100.0%; Pred. No. 3.8e-151;
Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60
Db 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60

Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120
Db 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120

Qy 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
Db 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180

Qy 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240
Db 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240

Qy 241 GTCTATGAGCGGTTTCGATATTAAAGTCCAGGATTAATTAAGTAAAGTAAAGTAAAGT 300
Db 241 GTCTATGAGCGGTTTCGATATTAAAGTCCAGGATTAATTAAGTAAAGTAAAGTAAAGT 300

Qy 301 GATTTCTAAACAGCTACTTATCTTGAATTACAGCATTTATGCTTATATGCTTCTGTTCT 360
Db 301 GATTTCTAAACAGCTACTTATCTTGAATTACAGCATTTATGCTTATATGCTTCTGTTCT 360

Qy 361 CGTGACATATGCTCTCGCGGTTTCTCTAGCCCTATCTGTTATATCCGATGTCGATCC 420
Db 361 CGTGACATATGCTCTCGCGGTTTCTCTAGCCCTATCTGTTATATCCGATGTCGATCC 420

Qy 421 TGATCAACTTTGGAATAAATCGGACGAGCTGAAATTTGATATGATATGATATGATATGAT 480
Db 421 TGATCAACTTTGGAATAAATCGGACGAGCTGAAATTTGATATGATATGATATGATATGAT 480

Qy 481 TTTAAGACAGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAATAAATAACAA 540
Db 481 TTTAAGACAGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAATAAATAACAA 540

Qy 541 TATTGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 600
Db 541 TATTGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 600

Qy 601 TATGATGGAACAAATACACATATATCAATAGTCTGCTAAAGAAATTCGAAATCGGTTT 660
Db 601 TATGATGGAACAAATACACATATATCAATAGTCTGCTAAAGAAATTCGAAATCGGTTT 660

Qy 661 AGCCAAACCAATCAGCATTTGTTAGTGCACCAATGTTAGTGCACCAATGTTAGTGCAC 720
Db 661 AGCCAAACCAATCAGCATTTGTTAGTGCACCAATGTTAGTGCACCAATGTTAGTGCAC 720

Qy 721 TTTCTGCTGCGGTAGGAGGTTATAGAGATAAATGCAATTTGGTTCGCTCACG 780
Db 721 TTTCTGCTGCGGTAGGAGGTTATAGAGATAAATGCAATTTGGTTCGCTCACG 780

Qy 781 CATTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTTCAATACCTFACATGCGGCAT 840
Db 781 CATTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTTCAATACCTFACATGCGGCAT 840

Qy 841 GTCTATGAGTCTTCTGTTGTTATGAATTCCTAA 874
Db 841 GTCTATGAGTCTTCTGTTGTTATGAATTCCTAA 874
```

```
RESULT 4
AF187007
LOCUS      874 bp      DNA      linear      BCT 07-MAR-2000
DEFINITION Haemophilus ducreyi strain CIP542 (CDC) serum resistance protein
            Dsra (dsrA) gene, complete cds.
ACCESSION AF187007
VERSION    AF187007.1 GI:7188584
KEYWORDS
SOURCE     Haemophilus ducreyi
ORGANISM   Haemophilus ducreyi
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE  1 (bases 1 to 874)
            Elkins C., Morrow K.J. Jr. and Olsen, B.
            Serum resistance in Haemophilus ducreyi requires outer membrane
            protein Dsra
            Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE    20143779
PUBMED     10678980
REFERENCE  2 (bases 1 to 874)
            Elkins, C., Morrow, K.J. and Olsen, B.
            Direct Submission
            Submitted (17-SEP-1999) Division of Infectious Diseases, Department
            of Medicine, University of North Carolina at Chapel Hill, 521
            Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
FEATURES
            Location/Qualifiers
            1..874
                /organism="Haemophilus ducreyi"
                /mol_type="genomic DNA"
                /strain="CIP542 (CDC)"
                /db_xref="taxon:730"
                /note="serum resistant strain; virulent; obtained from
                CDC"
            101..874
                /gene="dsrA"
            101..874
                /gene="dsrA"
                /note="outer membrane protein required for serum
                resistance"
                /codon_start=1
                /transl_table=11
                /product="serum resistance protein DsrA"
                /protein_id="AAF37813.1"
                /db_xref="GI:7188585"
                /translation="MKIKCLVAVVGLACSTTTMAQOPKFAVSSLYSEYDYGKCK
                WTMSNEGDFIKVPKIKMKPKWISKOATYLELOHYMPYPLVTVYAGVSPFLLY
                PMSDDQGLNROQLKUNLYSYFNDLRHDFKLKVLDIRISKNQKNDITISKYLLLEGT
                YLDSSYRMQONTHNINKLSKELOQTGLANQSLMLVOPNGVKGTSVSAAGVGRDXT
                ALAIGVSRITDRFTAKAGVAFNTYNGMSYGASVGYEF"
ORIGIN
Query Match      74.8%; Score 874; DB 1; Length 874;
Best Local Similarity 100.0%; Pred. No. 3.8e-151;
Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60
Db 1 ATAAATACGTCATTCACATTTTATTAAGGTAGATAAGAAAGTAAATCTATATT 60

Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120
Db 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTAATGTTTAGT 120

Qy 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
Db 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180

Qy 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240
Db 181 TGCTGAGTATCTTCTTTGTATAGCTATGATATGATATGATATGATATGATATGATATG 240

Qy 241 GTCTATGAGCGGTTTCGATATTAAAGTCCAGGATTAATTAAGTAAAGTAAAGTAAAGT 300
Db 241 GTCTATGAGTCTTCTGTTGTTATGAATTCCTAA 874
```


Ds	241	GTCTAATGAAGCGGTTTCGATATATAAGTGCCAGGATTAATAAGTGAAGCCAAAGAATG	300
Qy	301	GATTCTTAAACAGGCTACTTATCTTGAATACAGATTTATATGCTTATATCTCTGTTCT	360
Ds	301		
Ds	301	GATTCTTAAACAGGCTACTTATCTTGAATACAGATTTATATGCTTATATCTCTGTTCT	360
Qy	361	CGTGACATATGCTCCTGGCGTTCTCTAGCCCTATATCTGTATATCCGATGTCGATCC	420
Ds	361	CGTGACATATGCTCCTGGCGTTCTCTAGCCCTATATCTGTATATCCGATGTCGATCC	420
Qy	421	TGATCAACTTGGATAAATCGGAGCAGCTGAATTTGAATTTGTATATTTTAAACGA	480
Ds	421		
Qy	481	TTTAAGCACCATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAATAAACAAA	540
Ds	481	TTTAAGCACCATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAATAAACAAA	540
Qy	541	TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG	600
Ds	541	TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG	600
Qy	601	TATGATGAAACAAAATACACATATCAATTAAGTTGTCTAAAGAAATTCGAACTGGTTT	660
Ds	601	TATGATGAAACAAAATACACATATCAATTAAGTTGTCTAAAGAAATTCGAACTGGTTT	660
Qy	661	AGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT	720
Ds	661	AGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT	720
Qy	721	TTCTGCTGCGTAGGAGGTTATAGAGATPAAATCGCATTTAGCCATTTGGTGTGCGCTCACG	780
Ds	721	TTCTGCTGCGTAGGAGGTTATAGAGATPAAATCGCATTTAGCCATTTGGTGTGCGCTCACG	780
Qy	781	CATTACTGATCGCTTTACCGCTAAAGCGGTTGATGATTTCAATACCTACAAATGCGGCGAT	840
Ds	781	CATTACTGATCGCTTTACCGCTAAAGCGGTTGATGATTTCAATACCTACAAATGCGGCGAT	840
Qy	841	GTCTTATGCTCTCTGTTGGTTATGAATTTCTAA	874
Ds	841	GTCTTATGCTCTCTGTTGGTTATGAATTTCTAA	874
RESULT 5			
LOCUS	AF187004	847 bp	DNA linear BCT 07-MAR-2000
DEFINITION	Haemophilus ducreyi strain CIP542 (Can) serum resistance protein		
ACCESSION	DsrA (dsrA) gene, complete cds.		
VERSION	AF187004		
KEYWORDS	AF187004.1 GI:7188578		
SOURCE	Haemophilus ducreyi		
ORGANISM	Haemophilus ducreyi		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;		
AUTHORS	1 (bases 1 to 847)		
TITLE	Elkins, C., Morrow, K. J., Jr. and Olsen, B.		
JOURNAL	Serum resistance in Haemophilus ducreyi requires outer membrane		
MEDLINE	protein DsrA		
PUBMED	Infect. Immun. 68 (3), 1608-1619 (2000)		
AUTHORS	2 (bases 1 to 847)		
JOURNAL	Elkins, C., Morrow, K. J. and Olsen, B.		
TITLE	Direct Submission		
FEATURES	Submitted (17-SEP-1999) Division of Infectious Diseases, Department		
source	of Medicine, University of North Carolina at Chapel Hill, 521		
	Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA		
	Location/Qualifiers		
	1...847		
	/organism="Haemophilus ducreyi"		
	/mol_type="genomic DNA"		
	/strain="CIP542 (Can)"		

gene	/db xref="taxon:730"			
CDS	/notes="serum susceptible strain; obtained from William Albritton"			
	101..847			
	/genes="dsrA"			
	101..847			
	/genes="dsrA"			
	/notes="outer membrane protein required for serum resistance; deletion within coding region results in premature stop codon"			
	/codon_start=1			
	/transl_table=11			
	/product="serum resistance protein DsrA"			
	/protein_id="AAF37810.1"			
	/db xref="GI:7188579"			
	/translations="MKIKLVAVVGLACSTTTMAQPPKPFAGVSSILSYEYDYGKGK			
	WTWSNGGDDIKVPGIKMKPKWISQATYLEQHYMPYTPVLVYAPGVSPPILLY			
	PMSDDPOLGINRQOLKINLYSYFNDFDKLVDARISKNONIDITISKYLLLELGT			
	YLDSDYRMEQNTHNINKLSKELQTLGANSALSMVLQPNVGKTSYSAAVGGYRDKT			
	ALAIGVGSRIITDRFTAKAGVAFNTFYRNF"			
ORIGIN				
Query Match	71.0%	Score 829.2;	DB 1;	Length 847;
Best Local Similarity	99.6%	Pred. No. 6.8e-143;		
Matches	831;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
Qy	1	ATAAATACGTCATTGACATTTTATGTAAGGTAGATAAAGAACTAAATTTCTATATT	60	
Ds	1	ATAAATACGTCATTGACATTTTATGTAAGGTAGATAAAGAACTAAATTTCTATATT	60	
Qy	61	TACAATCAAGATTGACAAATTTATCTTAATGAGGTGATTTGAAAAATTAATTTAGT	120	
Ds	61	TACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTTGAAAAATTAATTTAGT	120	
Qy	121	TGCGTAGTGGGATTAGCTTGTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT	180	
Ds	121	TGCGTAGTGGGATTAGCTTGTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT	180	
Qy	181	TGCTGGAGTATCTTTTGTATAGCTATGATGATGACTATGGTAAGGTTAAATGGACTTG	240	
Ds	181	TGCTGGAGTATCTTTTGTATAGCTATGATGATGACTATGGTAAGGTTAAATGGACTTG	240	
Qy	241	GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAATAAGCCAAAGAATG	300	
Ds	241	GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAATAAGCCAAAGAATG	300	
Qy	301	GATTTCTAAACAGCTACTTATCTTGAATACAGCATTTATGCTTATCTCTGTTCT	360	
Ds	301	GATTTCTAAACAGCTACTTATCTTGAATACAGCATTTATGCTTATCTCTGTTCT	360	
Qy	361	CGTGACATATGCTCTGGCGTTTCTCCTAGCCCTATATCTGTTATATCCGATGTCGATCC	420	
Ds	361	CGTGACATATGCTCTGGCGTTTCTCCTAGCCCTATATCTGTTATATCCGATGTCGATCC	420	
Qy	421	TGATCAACTTGGATAAATCGGAGCAGCTGAAATTTGTAATTTGTATAGTTTAAACGA	480	
Ds	421	TGATCAACTTGGATAAATCGGAGCAGCTGAAATTTGTAATTTGTATAGTTTAAACGA	480	
Qy	481	TTTAAGCACGATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAATAAACAAA	540	
Ds	481	TTTAAGCACGATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAATAAACAAA	540	
Qy	541	TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG	600	
Ds	541	TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG	600	
Qy	601	TATGATGAAACAAAATACACATATCAATTAAGTTGTCTAAAGAAATTCGAACTGGTTT	660	
Ds	601	TATGATGAAACAAAATACACATATCAATTAAGTTGTCTAAAGAAATTCGAACTGGTTT	660	
Qy	661	AGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT	720	
Ds	661	AGCCAAACCAATCAGCATTTGCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT	720	


```

Qy 721 TTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCAATTAGCCATTGGTGTGCGCTCACG 780
Db 721 TTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCAATTAGCCATTGGTGTGCGCTCACG 780
Qy 781 CATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGCTTCAATACCTCAACAATGG 834
Db 781 CATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGCTTCAATACCTTCTATCG 834

RESULT 6
AF187002 890 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain CIPA75 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187002
VERSION AF187002.1 GI:7188574
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein DsrA.
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department
of Medicine, University of North Carolina at Chapel Hill, 521
Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
FEATURES
source
1..890
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CIPA75"
/db_xref="taxon:730"
/notes="serum susceptible strain"
96..890
/genes="dsrA"
96..890
/genes="dsrA"
/notes="outer membrane protein required for serum
resistance; no expression of DsrA in this strain due to a
5 bp deletion between putative -10 and -35 promoter
regions"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37808.1"
/db_xref="GI:7188575"
/translations="MKIKLVAVVGLACSTTTTAAQPPKPGVSSLYEYDYGK
WTWSNEGDFIKVPKIMKPEWISKQATLEIQLHYMPYPLVLTVAHDVPPSSILLY
PMSDPLQGLNKLNLKLYSYFNLDLRHDFKLVLDARISKQNTDITSKYLELGT
YLDSSYRMMEQNTNINKNTNINKLSKEIQTGLANQSALSLVQENGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGSVGEF"

Query Match 69.9%; Score 816.6; DB 1; Length 890;
Best Local Similarity 96.6%; Pred. No. 1.4e-140;
Matches 865; Conservative 0; Mismatches 4; Indels 26; Gaps 2;

Qy 1 ATAAATACGTCATTGACATTTTATGTAAGTACAATAAGAAAGTAATCTATATT 60
Db 1 ATAAATACGTCATTGACATTTT-----TAAGGTAGATAAAGAAAGTAATCTATATT 55
Qy 61 TACAATCAAGATTGACAAATTATTACTTAATGAGGTGATTATGAATAATTAATGTTTAGT 120
Db 56 TACAATCAAGATTGACAAATTATTACTTAATGAGGTGATTATGAATAATTAATGTTTAGT 115

```

```

Qy 121 TGCCGTAGTGGGANTAGCTTGTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
Db 116 TGCCGTAGTGGGANTAGCTTGTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 175
Qy 181 TGCTGAGATATCTTCTTTGTATAGCTATGATGATGACTATGGTAAAGGTAATGGACTTG 240
Db 176 TGCTGAGATATCTTCTTTGTATAGCTATGATGATGACTATGGTAAATGGACTTG 235
Qy 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTTAAATGAAGCAAAAGATG 300
Db 236 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTTAAATGAAGCAAAAGATG 295
Qy 301 GATTTCTTAACAGCTACTTATCTTGAATTTACAGCATTATGCCCTTATCTCTGTCT 360
Db 296 GATTTCTTAACAGCTACTTATCTTGAATTTACAGCATTATGCCCTTATCTCTGTCT 355
Qy 361 CGTGACATATGCTTCCTGGGCTTCTCTAGCCCTTATCTAGCTTATATCCGATGTCGATCC 420
Db 356 CGTGACATATGCTTCATGACGTTCTCTAGCTTATATCCGATGTCGATCC 415
Qy 421 TGATCAACTTGGAAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTAAACA 480
Db 416 TGATCAACTTGGAAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTAAACA 475
Qy 481 TTTAAGCACAGCTTTTAAATTTAAATTTAAAGTCTTGTATGACGCTATTTCCAAAATAAACA 540
Db 476 TTTAAGCACAGCTTTTAAATTTAAATTTAAAGTCTTGTATGACGCTATTTCCAAAATAAACA 535
Qy 541 TATTGATCTATATAAGTAAATATTATTAGAACTGGGTACTTTATTAGATGATCTTATTCG 600
Db 536 TATTGATCTATATAAGTAAATATTATTAGAACTGGGTACTTTATTAGATGATCTTATTCG 595
Qy 601 TATGATGGAACT-----AAAATACACATAATATCAATAAATAATCAATAAATTAAGTTGTC 639
Db 596 TATGATGGAACTAAATATACATAATATCAATAAATAATCAATAAATTAAGTTGTC 655
Qy 640 TAAAGATTTGCAAACTGTTTAGCCAACTCAATCAGCATTTCTCTAGTGTAGTCAACAAA 699
Db 656 TAAAGATTTGCAAACTGTTTAGCCAACTCAATCAGCATTTCTCTAGTGTAGTCAACAAA 715
Qy 700 TGGGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAATCTGCATT 759
Db 716 TGGGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAATCTGCATT 775
Qy 760 AGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTTACCGCTTAAAGCGGTGTAGCGTT 819
Db 776 AGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTTACCGCTTAAAGCGGTGTAGCGTT 835
Qy 820 CAATACCTACAATGGCGCATGCTTATGTTGCTTCTGTTGTTATGAATTTCTAA 874
Db 836 CAATACCTACAATGGCGCATGCTTATGTTGCTTCTGTTGTTATGAATTTCTAA 890

RESULT 7
AF187003 890 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain CIPA77 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187003
VERSION AF187003.1 GI:7188576
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980

```

REFERENCE 2 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..890
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CIPA77"
/db_xref="taxon:730"
/note="serum susceptible strain"
96..890
/gene="dsrA"
96..890
/gene="dsrA"
/note="outer membrane protein required for serum resistance; little or no expression of DsrA in this strain"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37809.1"
/db_xref="GI:7188577"

WTNSNEGFDIKVPGIKMKPKWKISQATYLELQHYNPYLPVLTYSADVPDPSSILLY
PMSDDQLGNRQOLKLNLYSYFNDLRHDFKLDARISKNKNIDITISKILLELGT
YLDGSYRMEQNTNKNTHNKNKLSKELQTLGANSALSMVQPNVGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGSASVGYEF"

ORIGIN
Query Match 69.9%; Score 816.6; DB 1; Length 890;
Best Local Similarity 96.6%; Pred. No. 1.4e-140;
Matches 865; Conservative 0; Mismatches 4; Indels 26; Gaps 2;

Qy 1 ATAAATACGTCATTGACATTTTATTAAGTAAAGTAGAATAAGAAAGTAAATCTATATT 60
Db 1 ATAAATACGTCATTGACATTTTATTAAGTAAAGTAGAATAAGAAAGTAAATCTATATT 55

Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAAATTAATGTTTAGT 120
Db 56 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAAATTAATGTTTAGT 115

Qy 121 TCGCGTAGTGGGATTAGTCTTCTACTATTACAAATGCGCTCAGCGCGCCAAAGTT 180
Db 116 TCGCGTAGTGGGATTAGTCTTCTACTATTACAAATGCGCTCAGCGCGCCAAAGTT 175

Qy 181 TCGTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTTAAATGCACTTG 240
Db 176 TCGTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTTAAATGCACTTG 235

Qy 241 GTCTAATGAAGCGCGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAATG 300
Db 236 GTCTAATGAAGCGCGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAATG 295

Qy 301 GATTCTTAAACAGGCTACTTATCTTGAATACAGATATATGCTTATACCTCGTTCT 360
Db 296 GATTCTTAAACAGGCTACTTATCTTGAATACAGATATATGCTTATACCTCGTTCT 355

Qy 361 CGTGACATATCTCTCGCGTTTCTCCPAGCCCTATCTGTTATATCCGATGTCGATCC 420
Db 356 CGTGACATATCTCTATGACGTTCTTCTTAGCTCTATACCTGTTATATCCGATGTCGATCC 415

Qy 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATATTTTAAACGA 480
Db 416 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATATTTTAAACGA 475

Qy 481 TTTAAGACACAGATTTTAAATTTAAAGTCTTGTATGTCAGGATTTCCAAAAATAACAAA 540
Db 476 TTTAAGACACAGATTTTAAATTTAAAGTCTTGTATGTCAGGATTTCCAAAAATAACAAA 535

Qy 541 TATTGATACTATAAGTAAATTTTACTAGAACTCGGCTACTTATTATTAGATGATCTTATCG 600

Db 536 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATCTTATCG 595
Qy 601 TATGATGGAC-----AAATACACATAATATCAATAGTTGTC 639
Db 596 TATGATGGACAAAATACACATAATATCAATAAAAATACACATAATATCAATAGTTGTC 655
Qy 640 TAAAGAAATGCAAACTGGTTTACCAACCAATCAGCATTCCTATGTTAGTGCAACAAA 699
Db 656 TAAAGAAATGCAAACTGGTTTACCAACCAATCAGCATTCCTATGTTAGTGCAACAAA 715
Qy 700 TGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATT 759
Db 716 TGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATT 775
Qy 760 AGCATTTGGTTCGGCTCAGCATTTACTGATCGCTTACCGCTAAACGGGTAGCGTT 819
Db 776 AGCATTTGGTTCGGCTCAGCATTTACTGATCGCTTACCGCTAAACGGGTAGCGTT 835
Qy 820 CAATACCTACAATGGCGCATGCTTATGCTGCTTCTGTTGGTTATCAATTTCTAA 874
Db 836 CAATACCTACAATGGCGCATGCTTATGCTGCTTCTGTTGGTTATCAATTTCTAA 890

RESULT 8
AF187009 916 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain Vii57 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187009
VERSION AF187009.1 GI:7188588
KEYWORDS Haemophilus ducreyi
SOURCE Haemophilus ducreyi
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 916)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 916)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..916
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="Vii57"
/db_xref="taxon:730"
/note="serum resistant strain"
101..916
/gene="dsrA"
101..916
/gene="dsrA"
/note="outer membrane protein required for serum resistance"
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37815.1"
/db_xref="GI:7188589"

WTNSNEGFDIKVPGIKMKPKWKISQATYLELQHYNPYLPVLTYSADVPDPSSILLY
PMSDDQLGNRQOLKLNLYSYFNDLRHDFKLDARISKNKNIDITISKILLELGT
YLDGSYRMEQNTNKNTHNKNKLSKELQTLGANSALSMVQPNVGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGSASVGYEF"

ORIGIN

Query Match		69.7%; Score 814; DB 1; Length 916;
Best Local Similarity		94.9%; Pred. No. 4.2e-140;
Matches 869; Conservative		0; Mismatches 5; Indels 42; Gaps 1;
QY	1	ATAAATACGTCATGTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATT 60
DB	1	ATAAATACGTCATGTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATT 60
QY	61	TACAATCAAGATTGACAAATTTATTTACTTAATAGAGTGATTAATGAAATTAATGTTAGT 120
DB	61	TACAATCAAGATTGACAAATTTATTTACTTAATAGAGTGATTAATGAAATTAATGTTAGT 120
QY	121	TGCGTAGTGGGATTAGTCTTCTACTATTACAATGGCTCAGCAGCGCCAAAGTT 180
DB	121	TGCGTAGTGGGATTAGTCTTCTACTATTACAATGGCTCAGCAGCGCCAAAGTT 180
QY	181	TGCTGGAGTATCTTTTGTATAGCTATGATGATGACTATGTAAGGTAAATGGAAGTTG 240
DB	181	TGCTGGAGTATCTTTTGTATAGCTATGATGATGACTATGTAAGGTAAATGGAAGTTG 240
QY	241	GTCTAATGAAGCGGTTTCATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 300
DB	241	GTCTAATGAAGCGGTTTCATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 300
QY	301	GATTTCTAAACAGGCTACTTATCTTGAATTAAGCATTATATGCTTATATCTCTGTTCT 360
DB	301	GATTTCTAAACAGGCTACTTATCTTGAATTAAGCATTATATGCTTATATCTCTGTTCT 360
QY	361	CGTGACATATGCTCTGGGCTTCTCTAGCCCTATCTATCTGTTATCCGATGCTGATCC 420
DB	361	CGTGACATATGCTCTGGGCTTCTCTAGCCCTATCTATCTGTTATCCGATGCTGATCC 420
QY	421	TGATCAACTTGGAAATTAATCGGAGCAGCTGAAATTTGATGATTTATTTAAACGA 480
DB	421	TGATCAACTTGGAAATTAATCGGAGCAGCTGAAATTTGATGATTTATTTAAACGA 480
QY	481	TTTAAGCAGCATTTTAAATTAAGTTCTGTATGACGATTTTCGAAATTAACAAATA 540
DB	481	TTTAAGCAGCATTTTAAATTAAGTTCTGTATGACGATTTTCGAAATTAACAAATA 540
QY	541	TATTGATCTATAGTAAATTTTACTAGAACCTGGGTACTTTTATGATGATCTTTATCG 600
DB	541	TATTGATCTATAGTAAATTTTACTAGAACCTGGGTACTTTTATGATGATCTTTATCG 600
QY	601	TATGATGGAAC-----AAATAC 618
DB	601	TATGATGGAACAAATACACATAATATCAATAAATAACATAATATCAATAAATAAC 660
QY	619	ACATAATATCAATAAGTTGCTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCAT 678
DB	661	ACATAATATCAATAAGTTGCTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCAT 720
QY	679	GTCCTATGTTAGTCACCAATGTTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGG 738
DB	721	GTCCTATGTTAGTCACCAATGTTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGG 780
QY	739	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCTCAGCATTAATGATCGCTTTAC 798
DB	781	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCTCAGCATTAATGATCGCTTTAC 840
QY	799	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGCGCATGTTTATGCGTCTCTGT 858
DB	841	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGCGCATGTTTATGCGTCTCTGT 900
QY	859	TGGTTATGAATTTCAA 874
DB	901	TGGTTATGAATTTCAA 916
RESULT 9		
AF187008		923 bp DNA linear BCT 07-MAR-2000
LOCUS		
DEFINITION		Haemophilus ducreyi strain M90-02 serum resistance protein DsrA (dsrA) gene, complete cds.
ACCESSION		AF187008
VERSION		AF187008.1 GI:7188586
KEYWORDS		
SOURCE		Haemophilus ducreyi
ORGANISM		Haemophilus ducreyi
REFERENCE		Bacteria: Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
AUTHORS		1 (bases 1 to 923)
TITLE		Elkins, C., Morrow, K.J. Jr. and Olsen, B.
JOURNAL		Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA
MEDLINE		Infect. Immun. 68 (3), 1608-1619 (2000)
PUBMED		20143779
REFERENCE		10678980
AUTHORS		2 (bases 1 to 923)
TITLE		Elkins, C., Morrow, K.J. and Olsen, B.
JOURNAL		Direct Submission
FEATURES		Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
source		Location/Qualifiers
1..923		/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"		
/strain="M90-02"		
/db_xref="taxon:730"		
/note="serum resistant strain"		
102..923		/gene="dsrA"
102..923		/gene="dsrA"
/note="outer membrane protein required for serum resistance"		
/codon_start=1		
/transl_table=11		
/product="serum resistance protein DsrA"		
/protein_id="AAP37814.1"		
/db_xref="GI:7188587"		
/translation="MKIKLVAVVGLACSTTTTAAQPPKFAVSSLYSEYDYGKGV		
WTWSEGGFDIKVPGIKMKPKWISQATYLEQHYMPYTPVLVTSAPDVSPSSISIL		
LYPMSPDQDGINROOLKLNLYSYFNDRHDFKLKVLDAKISKNKNIOTISKVLEL		
GYLDGSYRMMEQNTNINKNTHNINKLSKELQTLGANSLSMLVQPNQGV		
GKTSVSAVGGYRDKTALAIGVSRITDRFTAKGVAFNTNGMSYGASVGEF"		
ORIGIN		
Query Match		67.9%; Score 792.8; DB 1; Length 923;
Best Local Similarity		94.0%; Pred. No. 3.4e-136;
Matches 865; Conservative		0; Mismatches 7; Indels 48; Gaps 2;
QY	3	AAATACGTCATTTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATT 62
DB	4	AATACGTCATTTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATT 63
QY	63	CAATCAAGATTGACAAATTTATTTCTTAATGAGGTGATTAATAAATTAATGTTAGTTG 122
DB	64	CAATCAAGATTGACAAATTTATTTCTTAATGAGGTGATTAATAAATTAATGTTAGTTG 123
QY	123	CCGTAGTGGGATTAGCTTGTCTACTATTACAACAAATGGCTCAGCAGCGCCAAAGTTG 182
DB	124	CCGTAGTGGGATTAGCTTGTCTACTATTACAACAAATGGCTCAGCAGCGCCAAAGTTG 183
QY	183	CTGGAGTATCTCTTTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATG 242
DB	184	CTGGAGTATCTCTTTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATG 243
QY	243	CTAATGAAGCGGTTTCGATATTAAAGTGCAGGATTAATAATGAAGCCAAAGAATGGA 302
DB	244	CTAATGAAGCGGTTTCGATATTAAAGTGCAGGATTAATAATGAAGCCAAAGAATGGA 303
QY	303	TTTCTAAACAGGCTACTTCTTGAATTAACAGATTATATGCTTACTTCTGTTCTCG 362

```
Db 304 TTTCTAACACGGCTACTTATCTTGAATACAGCAATATATGCTCTATATGCTCTCTGCTCTCG 363
Qy 363 TGACATATGCTCTCGGCGTTCTCTAG-----CCCTATAGCTGTATATCCGATGCTG 416
Db 364 TGACATCTGCTCTGAGGTTCTCTAGCTCTATCTCTATAGCTGTATATCCGATGCTG 423
Qy 417 ATCTGTATCAACTTGGGAATAATCGGACGAGCTGAAATGAAATTCGTATGTTATTTTA 476
Db 424 ATCTGTATCAACTTGGGAATAATCGGACGAGCTGAAATGAAATTCGTATGTTATTTTA 483
Qy 477 ACGATTTAAGACAGATTTTAAATTAAGATTCGTATGACGATTTTCCAAAATAAAC 536
Db 484 ACGATTTAAGACAGATTTTAAATTAAGATTCGTATGACGATTTTCCAAAATAAAC 543
Qy 537 AAAATATTGATACATATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGATCTT 596
Db 544 AAAATATTGATACATATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGATCTT 603
Qy 597 ATCGTATGATGAAC-----AAA 614
Db 604 ATCGTATGATGAACAAAATAACATAATATCAATAAAAAATACACATAATATCAATAAAA 663
Qy 615 ATACACATAATATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTAGCCCAACCAATCAG 674
Db 664 ATACACATAATATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTAGCCCAACCAATCAG 723
Qy 675 CATTTGCTATGTTAGTGAACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTCGGTAG 734
Db 724 CATTTGCTATGTTAGTGAACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTCGGTAG 783
Qy 735 GAGGTTATAGAGATAAACTGCAATTAGCCATTTGTTGCGGTCAACGATTAATCTGATCGCT 794
Db 784 GAGGTTATAGAGATAAACTGCAATTAGCCATTTGTTGCGGTCAACGATTAATCTGATCGCT 843
Qy 795 TTACCGCTAAAGCGGTGTAGCGTTCAATACCTAACAATGCGGCGATGTTATGGTGCTT 854
Db 844 TTACCGCTAAAGCGGTGTAGCGTTCAATACCTAACAATGCGGCGATGTTATGGTGCTT 903
Qy 855 CTGTTGGTTATGAATTTCTAA 874
Db 904 CTGTTGGTTATGAATTTCTAA 923

RESULT 10
AF187006
LOCUS
DEFINITION Haemophilus ducreyi strain CH1A serum resistance protein Dsra
(dsrA) gene, complete cds.
ACCESSION AF187006
VERSION AF187006.1
KEYWORDS GI:7188582
SOURCE
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE
1 (bases 1 to 894)
Elkins, C., Morrow, K. J. Jr. and Olsen, B.
AUTHORS Serum resistance in Haemophilus ducreyi requires outer membrane
protein Dsra
TITLE Infect. Immun. 68 (3), 1608-1619 (2000)
JOURNAL
MEDLINE 20143779
PUBMED 10678980
REFERENCE
2 (bases 1 to 894)
Elkins, C., Morrow, K. J. and Olsen, B.
AUTHORS Direct Submission
TITLE Submitted (17-SEP-1999) Division of Infectious Diseases, Department
of Medicine, University of North Carolina at Chapel Hill, 521
Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
JOURNAL
FEATURES
Location/Qualifiers
1..894
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CH1A"
```

```
/db_xref="taxon:730"
/notes="intermediate serum susceptibility strain"
103..894
/gene="dsrA"
103..894
/gene="dsrA"
/notes="outer membrane protein required for serum
resistance"
/codon_start=1
/transl_table=11
/product="serum resistance protein Dsra"
/protein_id="AAF37812.1"
/db_xref="GI:7188583"
/translation="MKIKCLVAVVGLACSTITTTMAQPPKPFAGVSSLDSEYEDYKGGK
WWSKXGDGPDIKAPGIMKPKWISQATVGLQHYMPYFVLVTVASBPNTVLLYP
MPDPDOLGNROQLKLNLYSPNDLRHGFKLVDARI SONKONIDITISYLLKLGTV
LQSSYRMQNTNHNKNTNINKLSKELOTGLANQSALSMLVQPNVGKTSVSAAG
GYRDKTALAIGVGSRIIDRPTAKAGVAFNTNGMSYGSVGYEP"

ORIGIN
Query Match 66.8%; Score 780.6; DB 1; Length 894;
Best Local Similarity 94.1%; Pred. No. 6e-134;
Matches 842; Conservative 0; Mismatches 29; Indels 24; Gaps 2;
Qy 1 ATAAATACGTCATTCGACATTTTAAATGTAAGGTAGAATAAGAACTTAATTCATATT 60
Db 3 ATAAATACGTCATTCGACATTTTAAATGTAAGGTAGAATAAGAACTTAATTCATATT 62
Qy 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAATTTAAATGTTAGT 120
Db 63 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAATTTAAATGTTAGT 122
Qy 121 TSCCGTAGTGGGATTAGCTTGTCTACTATTACAACTGCTCAGCAGCGCCCAAGTT 180
Db 123 TSCCGTAGTGGGATTAGCTTGTCTACTATTACAACTGCTCAGCAGCGCCCAAGTT 182
Qy 181 TGCTGGAGTATCTTCTTTGTATAGTATGACTATGTAAGGGTAAATGGAGCTG 240
Db 183 TGCTGGAGTATCTTCTTTGGATAGTATGACTATGTAAGGGTAAATGGAGCTG 242
Qy 241 GTCTAATGAAGCGGTTTCGATATTAAGTGCAGGGATTAATAAGCAAGCAAGATG 300
Db 243 GTCTGAAGAAAGACGGTTTCGATATTAAGCGCCAGGGATTAATAAGCAAGCAAGATG 302
Qy 301 GATTTCTAAGACAGCTACTTATCTTGAATTAAGTGCAGCATTAATGCTCTTATCTCTGTTCT 360
Db 303 GATTTCTAAGACAGCTACTTATCTTGAATTAAGTGCAGCATTAATGCTCTTATCTCTGTTCT 362
Qy 361 CGTGACATATGCTCTCTGGCGTTTCTCTAGCCCTTATCTGTTATATCCGATGTCGATCC 420
Db 363 CGTGACATATGCTCTCTGCGAG--AACCTAACACTGTACTGTATATCCGATGTCGATCC 419
Qy 421 TGATCAACTTGGAAATAATCGGACGAGCTGAAATTTGAATTTGTTATAGTTATTTTAACGA 480
Db 420 TGATCAACTTGGAAATAATCGGACGAGCTGAAATTTGTTATAGTTATTTTAACGA 479
Qy 481 TTTAAGACAGATTTTAAATTTAAAGTCTTGTATGACGATTTTCCAAAATAAACAAA 540
Db 480 TTTAAGACAGATTTTAAATTTAAAGTCTTGTATGACGATTTTCCAAAATAAACAAA 539
Qy 541 TATTGATCTATTAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATGTTCTTATCG 600
Db 540 TATTGATCTATTAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATGTTCTTATCG 599
Qy 601 TATTGATGGAAC-----AAAATACATATATCAATAAGTTGTC 639
Db 600 TATTGATGGAACAAAATAACATAATATCAATAAAAATACATATATCAATAAGTTGTC 659
Qy 640 TAAAGAAATTCAAAACCTGTTTAGCCCAACCAATTCGCTATGTTAGTGCACCAAAA 699
Db 660 TAAAGAAATTCAAAACCTGTTTAGCCCAACCAATTCGCTATGTTAGTGCACCAAAA 719
Qy 700 TGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGATAAAACTGCATT 759
```

```

Db      720 TGGTGTAGGCACAAACGAGCGTTTCTGCTGCGGTAGGAGGTTATAGAGATAAACTGCATT 779
Qy      760 AGCCATTGGTGTGCGCTCAGCATTACTGATTCGCTTTACCGCTAAACGGGTGAGCGTT 819
Db      780 AGCCATTGGTGTGCGCTCAGCATTACTGATTCGCTTTACCGCTAAACGGGTGAGCGTT 839
Qy      820 CAATACCTACAAATGGCGGCATGCTTATGCTGCTTCTGTTGTTATGAATTTCTAA 874
Db      840 CAATACCTACAAATGGCGGCATGCTTATGCTGCTTCTGTTGTTATGAATTTCTAA 894

RESULT 11
AB064943 1967 bp DNA linear BCT 24-MAY-2002
LOCUS
DEFINITION Actinobacillus actinomycetemcomitans omp100 gene for outer membrane
protein 100, complete cds.
ACCESSION AB064943
VERSION
KEYWORDS
SOURCE
ORGANISM
Actinobacillus actinomycetemcomitans
Actinobacillus actinomycetemcomitans
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
REFERENCE
1 Komatsuzawa,H., Asakawa,R., Kawai,T., Ochiai,K., Fujiwara,T.,
Taubman,M.A., Ohara,M., Kurihara,H. and Sugai,M.
Identification of six major outer membrane proteins from
Actinobacillus actinomycetemcomitans
Gene 288 (1-2), 195-201 (2002)
MEDLINE 22030468
PUBMED 12034509
REFERENCE
2 (bases 1 to 1967)
Komatsuzawa,H.
Direct Submission
TITLE Submitted (04-JUL-2001) Hitoshi Komatsuzawa, Hiroshima University,
Faculty of Dentistry, Department of Microbiology; Minami-ku, Kasumi
1-2-3, Hiroshima 734-8553, Japan
(B-mail:hkomatsu@hiroshima-u.ac.jp, Tel:81-82-257-5637,
Fax:81-82-25-75639)
FEATURES
source
1. .1967
Location/Qualifiers
/organism="Actinobacillus actinomycetemcomitans"
/mol_type="genomic DNA"
/db_xref="taxon:714"
602..1489
/gene="omp100"
602..1489
/gene="omp100"
/codon_start=1
/transl_table=11
/product="outer membrane protein 100"
/protein_id="BAB86905.1"
/db_xref="GI:19911154"
/translation="MTYQLPKHLVLMVTGALSIVNALAKDSFLENPSANLPQVFN
RVDIFNNETNNKKDIANKIANISIEKDVNRNTGGIDRLAKQELVNRKITKNEL
DIRKNTKSTARIENASTARIDNGEVNRVLQNDVDRSTENARSRANEQKIAENKAI
ENKADKADVEKNKADIAANSRAIRATFRSSQNTAALTTKVDRNTARIDRLDSRVNLD
KEVKNGLASQALSGLPQPVNVSGLNSLSAAVGVKSKTALAVGSYRFNQVAAKAGV
AVTNGSGSATYVNLNPEW"

Query Match 7.2%; Score 84.6; DB 1; Length 1967;
Best Local Similarity 51.5%; Pred. No. 7.8e-06;
Matches 195; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

Qy      497 AAATTAAGTTCTTGATGACGATTTTCCAAAATAAACAATAATATTGATCTATAGT 556
Db      1112 AAAACACAGACCGCATATTGACGAAATATCCAGACGATTGCAACCTTTAGATCTTCAAGC 1171
Qy      557 AAATATTACTAGACGGGTACTTATTTAGATGATCTTATCGTATGATGGAACAAAT 616
Db      1172 CAAAACATCGCGCATTAACGACCAAAAGTTGATCGTAACTACTGCGCGTATTGATCGATTA 1231

```

```

Qy      617 ACACATAATATCAATAGTTTGTCTAAGAAATTCGAACACTGTTTAGCCACCAATCAGCA 676
Db      1232 GATAGCGAGTCAATGAATTAGACAAGAAAGTAAAAACCGTTTGGCTTCCCAAGCGGCA 1291
Qy      677 TTCTCTATGTTAGTGCACCAACCAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGA 736
Db      1292 CTAAGCGGCTTATCCAAACGATATATGTCGGCAGCCTTAACCTTGAGTGAGCTGTTGGT 1351
Qy      737 GGTATTAGAGATAAACTGCAATTAGCCATTGGTGTGCGCTCAGCATTACTGATCGGTTT 796
Db      1352 GGTATAAATCTAAAAACAGCACTAGCGGTTGGTTCAGGTTATCGTTTCAATCAAAATGTA 1411
Qy      797 ACCGCTAAAGCGGTGAGGTTCAATACCTACAATGGCGGCATGCTTATGCTGCTTCT 856
Db      1412 GCCGCAAGCGCGGTGTGGCAGTAAGTACCAATGGTGGCAGCGCAACTATAAGTCGGT 1471
Qy      857 GTTGGTTATGAATTTCTAAAT 875
Db      1472 TTAACTTTGAGTGGTAAT 1490

RESULT 12
AF316502 2579 bp DNA linear BCT 21-MAR-2002
LOCUS
DEFINITION Actinobacillus actinomycetemcomitans putative topoisomerase IV
subunit B (Aa32-1-1) gene, partial cds; and putative
adhesin/invasin (Aa32-1-2) gene, complete cds.
ACCESSION AF316502
VERSION
KEYWORDS
SOURCE
ORGANISM
Actinobacillus actinomycetemcomitans
Actinobacillus actinomycetemcomitans
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
REFERENCE
1 (bases 1 to 2579)
Lepine,G., Li,L. and Ellen,R.P.
Cloning and characterization of three invasive genes of
Actinobacillus actinomycetemcomitans
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 2579)
Lepine,G., Li,L. and Ellen,R.P.
Direct Submission
TITLE Submitted (25-OCT-2000) Dental Research Institute, University of
Toronto, 124 Edward St., Room 442, Toronto, Ontario M5G 1G6, Canada
JOURNAL
LOCATION/Qualifiers
Location/Qualifiers
FEATURES
source
1. .2579
/organism="Actinobacillus actinomycetemcomitans"
/mol_type="genomic DNA"
/strain="UT32"
/db_xref="taxon:714"
complement (<1..882)
/gene="Aa32-1-1"
complement (<1..882)
/gene="Aa32-1-1"
/notes="Similar to Haemophilus influenzae topoisomerase IV
subunit B protein deposited in GenBank Accession Number
AAC23174"
/transl_table=11
/codon_start=1
/product="putative topoisomerase IV subunit B"
/protein_id="AAL91672.1"
/db_xref="GI:19568163"
/translation="MTNTYSAEITVLKLEPVLQRPQMYTDTTRPNHLAQEVIDNSV
DRALAGFATKIEVILHKDOSIEVIDNGRMPVDIHPVEKVGSEVIMSKLHAGKFSN
KNYTFSGGLHGVGISVUNALSERVDYTKRSEIYKIAPENGKKVDLTWIGTCGRPT
TGTVTFFKPNKFTDSEKFSKRLHLLKRAVLCVSGLEIKFKFDNDVNDTDTWLYQDG
LNDLYMEAVNGLVTFDPEQFPIGEFNGEKEAVSWALLPPEGSLIGESVYNLIPTALG
GTHVNLGRQLLDAMREF"
1385..2272
/gene="Aa32-1-2"
1385..2272
/gene="Aa32-1-2"

```

/note="similar to Neisseria meningitidis MC58
adhesin/invasin deposited in GenBank Accession Number
AAF42321 and Moraxella catarrhalis ubiquitous surface
protein A2 deposited in GenBank Accession Number AAB96391"
/codon_start=1
/transl_table=11
/product="putative adhesin/invasin"
/protein_id="AAL91673.1"
/db_xref="GI:19568164"
/translation="MTYOLPKHILVALMTVGAISVNALAKDSFLENPSANLPQOVFNQ
RVDIPNNETHNENKDIANKANIASIEKDVMTGCDRLAKQELVNRARTKNEL
DIQNTKSIASIAENTADGNLEGNVRVQNDVVRSTENAKSRANEQKIAENKKAI
ENKADADVENRADIARAIATFRSSQNTAALTTKVDRTARIDRLDSRVNELD
KEVNGLASQAALSGLPQPNVNSLNSAAVGGYKSTALAVSGSYRFNQVNAKAGV
AVSTNGGSATVNVGLNFEW"

ORIGIN

Query Match 7.2%; Score 84.6; DB 1; Length 2579;
Best Local Similarity 51.5%; Pred. No. 7.4e-06;
Matches 195; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

Qy 497 AAATTAAAGTCTTGATGACGATTTCCAAAATAAACAATAATTGACTATAAGT 556
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1895 AAAACAGAGCCGATATTGCAGCAATTCAGAGCGATTGCAACCTTTAGATCTCAAGC 1954

Qy 557 AAATATTACTAGAACTGGGTACTTATTATTGATGATCTTATCGTATGATGGAACAAAAT 616
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1955 CAAAACATCGCGGCATTAACGACCAAAAGTTGATCGTAATACTGCGCGTATTGATCGATTA 2014

Qy 617 ACACATATATCAATAGTTGCTTAAGAANTGCAAACTGGTTTAGCCAAACCAATCAGCA 676
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2015 GATAGCCGAGTCAATGAATTAGACAAAGAAAGTAAACCGTTTGGCTTCCCAAGCGCA 2074

Qy 677 TTGCTATGTTAGTGCACCAAAAGGTTAGTACGCAAAACGAGCGTTTCTGCTCGCGTAGGA 736
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2075 CTAGCGGCTTATTCCAAACCGTATAATGTCGCGACGCTTAACCTTGAGTGCAGCTGTTGGT 2134

Qy 737 GGTATTAGATAAACTGCAATGACCAATGGTGTGCGGCTCAGCCATTACTGATCGGCTTT 796
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2135 GGTATAAATCTAAACAGCACTAGCGGTGGTTTCAGGTATCGTTTCATCAAAATGTA 2194

Qy 797 ACCGCTAAGCGGTGTAGCGTTCAATACCTACAAATGCGCGCATGTCTTATGTCGCTTCT 856
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2195 GCCGGAAGGCGGTGTGCGAGTAGTACCAATGTTGCGACGCAACCTATAACGTCGGT 2254

Qy 857 GTTGGTTATGAATCTTAAT 875
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2255 TTAAACTTTGAGTGGTAAT 2273

RESULT 13
YPYOPA YPYOPA 2216 bp DNA linear BCT 12-SEP-1993
LOCUS Yersinia pseudotuberculosis virulence plasmid pIBI yopA gene for
DEFINITION YopI protein.
ACCESSION X13883 X12758
VERSION X13883.1 GI:48639
KEYWORDS invasin; plasmid; yopA gene.
SOURCE Yersinia pseudotuberculosis
ORGANISM Yersinia pseudotuberculosis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.
1 (bases 357 to 1729)
Rosqvist,R.; Skurnik,M. and Wolf-Watz,H.
Increased virulence of Yersinia pseudotuberculosis by two
independent mutations
Nature 334 (6182), 522-524 (1988)
JOURNAL 88302441
MEDLINE 3043229
PUBMED
REFERENCE 2 (bases 1 to 2216)
AUTHORS Skurnik,M. and Wolf-Watz,H.
TITLE Analysis of the yopA gene encoding the YopI virulence determinants
of Yersinia spp

Mol. Microbiol. 3 (4), 517-529 (1989)
JOURNAL 89343638
MEDLINE 2761389
PUBMED
REFERENCE 3 (bases 1 to 2216)
AUTHORS Skurnik,M.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1988) Skurnik M., Department of Medical
Microbiology, University of Turku, SF-20520 Turku, Finland
COMMENT Data kindly reviewed (05-JUL-1989) by Skurnik M.
FEATURES Location/Qualifiers
source 1..2216
/organism="Yersinia pseudotuberculosis"
/mol_type="genomic DNA"
/strain="YP111"
/db_xref="taxon:633"
/clone="pYMS 2, pIBI540"
/plasmid="pIBI"
54..82
promoter /notes="promoter region"
88..1720
precursor_RNA /notes="yopA transcript"
357..1661
CDS /note="unamed protein product; YopI preprotein (AA 1 -
434)"
/codon_start=1
/transl_table=11
/protein_id="CAA32088.1"
/db_xref="GI:48640"
/db_xref="GOA:P10858"
/db_xref="SWISS-PROT:P10858"
/translation="MTKDFKISVSAALISALFSPYAFABEEDGNDGI PRLSAVQIS
PNVDPKLVGLYPAKPIILROENPKLPRGPQPKKRLAEAIQPOVLGAGGLNARA
KDPIAIGARASATAAAPAAVAVGSGIATGVNSVAIGHSPKALGSDSAVTVGASSTAK
DGVIAIGARASATGVAVGNSKVDAGNSVAIGHSHVADHGYSIAIGHSKTDREN
SVSIGHESLNRLTHLAAGTEDTDAVNAQLKXEMAEITLENARKETLAQSDVLDAAK
KHNSVQYTTDHKFSOLDNRDLKRDVKGSLAASALNSLFPQYGVGVNFTAGVGGY
AISESNQYTDHKEFOLDNRDLKRDVKGSLAASALNSLFPQYGVGVNFTAGVGGY
RSSQALAIQSGSYRVNESVALKAGVAVAGSSNMVYNASFNIEW"
1695..1720
terminator /note="transcription terminator"

ORIGIN

Query Match 5.9%; Score 68.6; DB 1; Length 2216;
Best Local Similarity 54.6%; Pred. No. 0.0067;
Matches 137; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 568 AGAAGTGGTACTATTATGATGATCTTATCGTATGATGGAACAAATACACATAAT 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1358 AGCAATCAGTGAATCTAATCAATACACAGATCAATAATCAGTCAACTTGACACCGGTT 1417

Qy 628 CAATAAGTTGCTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCATGTCTATGTT 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1418 AGATAAATTGACAAACGAGTTGCAAAAGTTTAGCCAGTTTCAGCCGCTTTAAACAGCTT 1477

Qy 688 AGTGCAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGAGGTTATAGAGA 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1478 GTTCCAGCCATATGGTGTAGGAAAGTAAACTTTACTGCAAGTGTGCGGGGATATCGTTC 1537

Qy 748 TAAACTGCATTAGCCATTTGGTGTGCGCTCAGCATTTACTGATCGCTTACCGGTAAGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1538 TAGTCAGCATTAGCAATTTGCTTCTGCTATCGTGTAAATGAGAGTGTGCGCACTTAAAGC 1597

Qy 808 GGGTGTAGCCT 818
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1598 CGGTGTGCTT 1608

RESULT 14
YPYOPANF YPYOPANF 2216 bp DNA linear BCT 01-DEC-1992
LOCUS Yersinia pestis virulence plasmid pYV019 yopA pseudogene.
DEFINITION Yersinia pestis virulence plasmid pYV019 yopA pseudogene.
ACCESSION X13880 X12759
VERSION X13880.1 GI:48641

KEYWORDS plasmid; pseudogene; putative adhesin; yopA gene.

SOURCE Yersinia pestis

ORGANISM Yersinia pestis

REFERENCE 1 (bases 1 to 2216)

AUTHORS Rosqvist, R., Skurnik, M., and Wolf-Watz, H.

TITLE Increased virulence of Yersinia pseudotuberculosis by two independent mutations

JOURNAL Nature 334 (6182), 522-524 (1988)

MEDLINE 88302441

PUBMED 3043229

REFERENCE 2 (bases 1 to 2216)

AUTHORS Skurnik, M., and Wolf-Watz, H.

TITLE Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia spp

JOURNAL Mol. Microbiol. 3 (4), 517-529 (1989)

MEDLINE 89343638

PUBMED 2761389

REFERENCE 3 (bases 1 to 2216)

AUTHORS Skurnik, M.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-1988) Skurnik M., Department of Medical Microbiology, University of Turku, SF-20520 Turku, Finland

COMMENT see X13883 for functional Y. pseudotuberculosis yopA gene. Data kindly reviewed (05-jul-1989) by Skurnik M.

FEATURES

source

1. .2216

/organism="Yersinia pestis"

/mol_type="genomic DNA"

/strain="019"

/db_xref="taxon:632"

/clone="pYWS 1, pYWS 1080."

/plasmid="pYV019"

4. .32

/note="promoter region"

307

/note="yopA translation start (Y. pseudotuberculosis)"

539

/note="deletion site"

1643. .1669

/note="transcription terminator"

ORIGIN

Query Match 5.9%; Score 68.4; DB 1; Length 2216;

Best Local Similarity 56.8%; Pred. No. 0.0073;

Matches 126; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 597 ATCGTATGATGGACAAATACACATATATCATCAATAGTTGCTTAAGAAATGCAAACTG 656

Db 1336 ATCATAAATTCAGTCAACTTGACAAACCGTTTAGATAACTGACAAACGAGTTGACAAAG 1395

Qy 657 GTTTAGGCAACCAATCAGCATTTGCTATGTTAGTGCAACAAATGGTGAGGCAAAACGA 716

Db 1396 GTTTAGGCAAGTTACGCGCTTTAAACAGCTTTGTCAGCCATATGTTAGGGAAAGTAA 1455

Qy 717 GCGTTTCGTCGCGTAGGAGGTTATAGATAAACTGATGACCATCGGTTGCGGCT 776

Db 1456 ACTTTCCTCGAGGTCGCGGGGATATCGTTCTAGTCAGGCATTAAGCAATGGTTCTGCTT 1515

Qy 777 CAGCATTAATGATCGCTTTTACCGCTAAAGCGGTGTAGCGT 818

Db 1516 ATCGTGTAATGAGAGTGTGCGCATTTAAAGCGGTGTGCGCTT 1557

RESULT 15

LOCUS YPCD1 70305 bp DNA circular BCT 18-JUN-2003

DEFINITION Yersinia pestis plasmid pCD1.

ACCESSION AL117189

VERSION AL117189.1 GI:5832423

KEYWORDS chaperone; cytotoxic effector; IS100; IS1616; IS1617; lcr; low-calcium response; syc; targeted effector; transposase; type III

secretion; V antigen; virulence; ylp; yop; ysc.

Yersinia pestis

Yersinia pestis

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.

1 (bases 1 to 70305)

AUTHORS Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Farraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrall, B.G.

TITLE Genome sequence of Yersinia pestis, the causative agent of plague

JOURNAL Nature 413 (6855), 523-527 (2001)

MEDLINE 21470413

PUBMED 11586360

REFERENCE 2 (bases 1 to 70305)

AUTHORS James, K.D., Parkhill, J., Barrall, B.G. and Rajandream, M.A.

TITLE Direct Submission

JOURNAL Submitted (03-SEP-1999) Yersinia pestis sequencing project, The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk DNA supplied by Dr. Andrew Karlyshev and Prof. Brendan Wren, [3]. Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT

Notes:

Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES

source

1. .70305

/organism="Yersinia pestis"

/mol_type="genomic DNA"

/strain="CO-92 Biovar Orientalis"

/db_xref="taxon:632"

/plasmid="pCD1"

1. .1356

/note="IS100 element"

88. .1110

/gene="YPCD1.01"

88. .1110

/gene="YPCD1.01"

/note="YPCD1.01, probable transposase, len: 340 aa; putative insertion sequence IS100, identical to corresponding CDS from Y.pestis KIM5 pCD1 (EMBL:AF053946, AF074612) (340 aa), fasta scores; opt: 2328 z-score: 3808.9 E(): 0, 100.0% identity in 340 aa overlap. Similar to many others e.g. TRAO_ECOLI (EMBL:X14793), istA, E.coli transposase for insertion sequence element IS21 (390 aa) (33.1% identity in 329 aa overlap). Contains Pfam match to entry PF00239 recombinase, Site-specific recombinases, score 25.70, E-value 4.8e-06. Contains probable helix-turn-helix motif at aa 19-40 (Score 2045, +6.15 SD)"

/codon_start=1

/transl_table=11

/product="putative transposase"

/protein_id="CAB54878.1"

/db_xref="GI:5832424"

/db_xref="GOA:P74993"

/db_xref="SPTREMBL:P74993"

/translation="MVFETVMEIKLHKQMSRAIARELGRNIVKRYLQKSP PKYTPRPAVASLLDEYRIQRIADAHPIKATVIAREIQDVGKGGMTILRAFR SLSVQEQEPAPAFETEPGRQMDVWNGRSPFLVFVAVLGYSRLMYIEFTDNMR YDTLETCHRNAPFRFGVPREVLYDNWKTVLORDAYQTQHRFHPSLWQFGKMGFSPRLCRPFRQAQTKGKVERMYQYTENSFVPLMTFLRWGITVDVETANRHLRLWLDVA NQRKHETIQARPCDRWLEEQSQSMALPPEKKEYVDVLDENLVNFDKHLHPPLSIYDS FCRGVA"

112. .195

/gene="YPCD1.01"

/note="Pfam match to entry PF00239 recombinase, Site-specific recombinases, score 25.70, E-value 4.8e-06"

1110. .1889

misc_feature

112. .195

/gene="YPCD1.01"

/note="Pfam match to entry PF00239 recombinase, Site-specific recombinases, score 25.70, E-value 4.8e-06"

1110. .1889

gene

101	ATGAAATTAAATGTTTAGTTGCCGTAGTGGGATTAGCTTGTCTTACTATTACACAAGT	160
Qy		
1	MetLysIleLysCysLeuValaValValGlyLeuAlaCysSerThrIleThrMet	20
Db		
161	GCTCAGCAGCCCAAGTTGCTGGAGTATCTCTTTGATAGCTATGATGACTACT	220
Qy		
21	AlaGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluLysAspTyr	40
Db		
221	GGTAAGGGTAAATGGACTTGGTCTAAATGAAGCGCGTTTCGATATTAAAGTGCACGGATT	280
Qy		
41	GlyLysGlyLysTyrThrTyrSerAsnGluGlyGlyPheAspIleLysValProGlyIle	60
Db		
281	AAATGAAGCCCAAGAAGATGATTTCTAAACAGCGTACTTATCTTGAAATTCACAGATTAT	340
Qy		
61	LysMetLysProLysGluTyrPheSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr	80
Db		
341	ATGCGCTTATACTCTCTGTTCTCTCGTACATATGCTCTCGCGTTCCTCCCTAGCCCTATACGT	400
Qy		
81	MetProTyrThrProValLeuValThrSerAlaProAspValProProSerSerIleLeu	100
Db		
401	TTATATCCGATGTCGTGATCCGTGATCAACTCGGAATAAATCGGCGAGCTGAAATTTGAAT	460
Qy		
101	LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn	120
Db		
461	TTGTATAGTATTATTTAAACGATTTAAGACACGATTTTAAATTTAAAGTTCTTTGATGCACGT	520
Qy		
121	LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg	140
Db		
521	ATTTCACAAAATAACAAATATTGATACTATAAGTAAATATTTACTAGAACTGGGTACT	580
Qy		
141	IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr	160
Db		
581	TATTTAGATGATCTTATCGTATGATGAACAA-----613	
Qy		
161	TyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr	180
Db		
614	-----AATACACATAATATCAATAAGTTGTCTAAAGAAATTTGCCAAACTGGT	658
Qy		
181	HisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGlnThrGly	200
Db		
659	TTAGCCAAACAATCAGCATTTGCTATGTTAGTGCAACCAAAATGTGTAGGCAAAACAGC	718
Qy		
201	LeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLysThrSer	220
Db		
719	GTCTCTGCTGCGGTAGGAGGTTATAGATAAAACTGCATTAGCCATTGGCTGGGCTCA	778
Qy		

```
QY 101 ATGAATAAATGCTTTAGTTAGTCCGTAGTGGATTAGCTTGTCTTACTATTACAAATG 160
Db |||||
QY 1 MetLysIleLysCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Db |||||
QY 161 GCTCAGCAGCGCCAAAGATTTGCTGGAGTACTCTTTGTATAGCTATGACTATGACTAT 220
Db |||||
QY 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Db |||||
QY 221 GGTAAAGGTAATCGACTGTGCTTAATGAAGCGCGTTTCGATATTAAGTCCAGGATT 280
Db |||||
QY 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Db |||||
QY 281 AAAATGAAGCCAAAGATGATTTCTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db |||||
QY 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Db |||||
QY 341 ATGCTTATATCTCTGTTCTGCTGACATATGCTCTCGCGTGTCTCTCTAGC-----CCT 394
Db |||||
QY 81 MetProTyrThrProValLeuValThrSerAlaProAspValSerProSerSerIleSer 100
Db |||||
QY 395 ATACTGTATATCCGATGTCGATCCCTGATCAACTGGAATAAATCGCAGAGCTGAAA 454
Db |||||
QY 101 IleLeuLeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnLeuLys 120
Db |||||
QY 455 TTGAATTGTATAGTATTTTAAAGATTTAAGACACGATTTTAAATTTAAAGTTCTTGAT 514
Db |||||
QY 121 LeuAsnLeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAsp 140
Db |||||
QY 515 GCAGTATTTCCAAAATAAACAATAATATTGATATAAGTAAATATTACTAGAACTG 574
Db |||||
QY 141 AlaArgIleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeu 160
Db |||||
QY 575 GGTACTTATTAGATGATCTTATCGTATGATGAACAA----- 613
Db |||||
QY 161 GlyThrTyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLys 180
Db |||||
QY 614 -----ATAACATAATCAATCAATAGTTGCTTAAAGATTCGAA 652
Db |||||
QY 181 AsnThrHisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGln 200
Db |||||
QY 653 ACTGGTTTAGCCAACTAGCATCTGCTATGTTAGTGAACCAAAATGTTGAGGCAAA 712
Db |||||
QY 201 ThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLys 220
Db |||||
QY 713 ACGAGCGTTTCTGCTGCGTAGGAGTTATAGATATAAACTGCATTAGCGTTGTC 772
Db |||||
QY 221 ThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeuAlaIleGlyVal 240
Db |||||
QY 773 GGCTCAGCATTTACTGATCGTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTACAAT 832
Db |||||
QY 241 GlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsn 260
Db |||||
QY 833 GCGCGCATGCTTATGCTGCTTCGTTGTTGGTTGATGAATTC 871
Db |||||
QY 261 GlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 273
Db |||||
RESULT 8
ID AAB31708 standard; protein; 242 AA.
XX
AC AAB31708;
XX
XX 30-APR-2001 (first entry)
XX
XX Amino acid sequence of the DsrA protein from strain CIP542 (Can).
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
XX Haemophilus ducreyi.
XX
XX W0200104138-A1.
PN
```

```
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000WO-US018834.
XX
XX 09-JUL-1999; 99US-0143257P.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX (ELKI/) ELKINS C.
XX
XX Elkins C;
XX
XX WPI; 2001-138311/14.
XX
XX N-PSDB; AAF25265.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX resistance to the bacteria used to produce vaccines that induce immune
XX response against the bacteria in subject at risk of developing chancroid.
XX
XX Claim 7; Page 55; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
XX ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX polypeptide is used to produce vaccine compositions, which are useful for
XX inducing a protective immune response in a subject at risk of developing
XX Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX fragments or peptides can be used for screening libraries of compounds in
XX a variety of drug screening techniques. The proteins and peptides may be
XX used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
XX
XX The nucleic acids are useful for the preparation of DsrA proteins
XX
XX Sequence 242 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,7e-138 Length: 242
XX Score: 1259.00 Matches: 242
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 61.47% Indels: 0
XX DB: 4 Gaps: 0
XX
XX US-10-030-529A-1 (1-1168) x AAB31708 (1-242)
QY 101 ATCAAAATTAATGTTTAGTTGCGTAGTGGATTAGCTTGTCTTACTATTACAAATG 160
Db |||||
QY 1 MetLysIleLysCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Db |||||
QY 161 GCTCAGCAGCGCCAAAGATTTGCTGGAGTACTCTTTGTATAGCTATGACTATGACTAT 220
Db |||||
QY 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Db |||||
QY 221 GGTAAAGGTAATCGACTGTGCTTAATGAAGCGCGTTTCGATATTAAGTCCAGGATT 280
Db |||||
QY 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Db |||||
QY 281 AAAATGAAGCCAAAGATGATTTCTTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db |||||
QY 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Db |||||
QY 341 ATGCTTATATCTCTGTTCTGCTGACATATGCTCTCGCGTGTCTCTCTAGC-----CCT 394
Db |||||
QY 81 MetProTyrThrProValLeuValThrSerAlaProAspValSerProSerSerIleSer 100
Db |||||
QY 395 ATACTGTATATCCGATGTCGATCCCTGATCAACTGGAATAAATCGCAGAGCTGAAA 454
Db |||||
QY 101 IleLeuLeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnLeuLys 120
Db |||||
QY 455 TTGAATTGTATAGTATTTTAAAGATTTAAGACACGATTTTAAATTTAAAGTTCTTGAT 514
Db |||||
QY 121 LeuAsnLeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAsp 140
Db |||||
QY 515 GCAGTATTTCCAAAATAAACAATAATATTGATATAAGTAAATATTACTAGAACTG 574
Db |||||
QY 141 AlaArgIleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeu 160
Db |||||
QY 575 GGTACTTATTAGATGATCTTATCGTATGATGAACAA----- 613
Db |||||
QY 161 GlyThrTyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLys 180
Db |||||
QY 614 -----ATAACATAATCAATCAATAGTTGCTTAAAGATTCGAA 652
Db |||||
QY 181 AsnThrHisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGln 200
Db |||||
QY 653 ACTGGTTTAGCCAACTAGCATCTGCTATGTTAGTGAACCAAAATGTTGAGGCAAA 712
Db |||||
QY 201 ThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLys 220
Db |||||
QY 713 ACGAGCGTTTCTGCTGCGTAGGAGTTATAGATATAAACTGCATTAGCGTTGTC 772
Db |||||
QY 221 ThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeuAlaIleGlyVal 240
Db |||||
QY 773 GGCTCAGCATTTACTGATCGTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTACAAT 832
Db |||||
QY 241 GlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsn 260
Db |||||
QY 833 GCGCGCATGCTTATGCTGCTTCGTTGTTGGTTGATGAATTC 871
Db |||||
QY 261 GlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 273
Db |||||
RESULT 8
ID AAB31708 standard; protein; 242 AA.
XX
XX AAB31708;
XX
XX 30-APR-2001 (first entry)
XX
XX Amino acid sequence of the DsrA protein from strain CIP542 (Can).
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
XX Haemophilus ducreyi.
XX
XX W0200104138-A1.
PN
```

Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 Qy 521 ATTTCCAAAATAAACAAATATTGATACTATAAGTAAATATTTACTAGAACTGGGTACT 580
 Db 141 ILeSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
 Qy 581 TATTAGATGATCTTATCGTATGATGGAACAAATACACATAATATCAATAAGTTGTCT 640
 Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
 Qy 641 AAGAATGTGCAACTGGTTAGCCAAACCAATCAGCATGTCTATGTTAGTGCACCAACAAAT 700
 Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
 Qy 701 GGTGTAGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGATAAAACTGCATTA 760
 Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeu 220
 Qy 761 GCCATTGGTGGCTCAGCATTAAGTATGATCGCTTTACCGCTAAAGCGGGGTAGCGTTC 820
 Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
 Qy 821 AATACC 826
 Db 241 AsnThr 242
 RESULT 9
 AAB311710
 ID AAB311710 standard; protein; 263 AA.
 XX AC AAB311710;
 XX DT 30-APR-2001 (first entry)
 XX DE Amino acid sequence of the DsrA protein from strain CH1A.
 XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX KW chancroid disease; genital ulcer disease.
 XX OS Haemophilus ducreyi.
 XX PN WO200104138-A1.
 XX PD 18-JAN-2001.
 XX PF 07-JUL-2000; 2000WO-US018834.
 XX PR 09-JUL-1999; 99US-0143257P.
 XX PA (UYNK-) UNIV NORTH CAROLINA.
 XX PA (ELKI/) ELKINS C.
 XX PI Elkins C;
 XX DR WPI; 2001-138311/14.
 XX DR N-PSDB; AAF25267.
 XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX PS Claim 7; Page 56; 80pp; English.
 XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be

CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX SQ Sequence 263 AA;
 Alignment Scores:
 Pred. No.: 8,22e-132 Length: 263
 Score: 1205.00 Matches: 235
 Percent Similarity: 92.05% Conservative: 8
 Best Local Similarity: 89.02% Mismatches: 13
 Query Match: 58.84% Indels: 8
 DB: Gaps: 2
 US-10-030-529A-1 (1-1168) x AAB311710 (1-263)
 Qy 101 ATGAATAATTAATGTTAGTTGCGGTAGTGGGATTAGCTTGTCTACTATTACACAATG 160
 Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
 Qy 161 GCTCAGCAGCGCCCAAGTTTGTGGAGTATCTTCTTTGTATAGCTATGATGACTAT 220
 Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuAspSerTyrGluTyrAspTyr 40
 Qy 221 GGTAAAGGTAAATCGACTTCGTCTAATGAAGCGGGTTTCGATATATAAGTCCAGGATT 280
 Db 41 GlyLysGlyLysTrpThrTrpSerGluLysAspGlyPheAspIleLysAlaProGlyIle 60
 Qy 281 AAAATGAGCCAAAGAGATGATTTCTAAACAGGCTACTTATCTTGAATACAGCATAT 340
 Db 61 LysMetLysProlLysLysTrpIleSerArgGlnAlaThrTyrLeuGlyLeuGlnHisTyr 80
 Qy 341 ATGCTTATATCTCTGTTCTCGTACATATGCTCGCGTCTTCTCTAGCCCTATACTG 400
 Db 81 MetProTyrThrProValLeuValThrTyrAla---SerAlaGluProAsnThrValLeu 99
 Qy 401 TTATATCCGATGCTGATCTGATCAACTGGAATAAAATCGGACGAGCTGAAATTTGAAT 460
 Db 100 LeuTyrProMetProAspProAspGlnLeuGlyLeuAsnArgGlnGlnLeuLysLeuAsn 119
 Qy 461 TTGTATAGTTTATTAAACGATTTAAGACACGATTTAAATTTAAAGTCTTGTATGACAGT 520
 Db 120 LeuTyrSerTyrPheAsnAspLeuArgHisGlyPheLysLeuAsnValLeuAspAlaArg 139
 Qy 521 ATTTCCAAAATAAACAAATATTGATACTATAAGTAAATATTTACTAGAACTGGGTACT 580
 Db 140 ILeSerGlnAsnLysGlnAsnIleAspThrIleSerGluTyrLeuLeuLysLeuGlyThr 159
 Qy 581 TATTAGATGATCTTATCGTATGATGGAACAA-----AATACA 619
 Db 160 TyrLeuAspSerSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 179
 Qy 620 CATATATCAATAAGTTGTCTTAAGAATTCGAACCTGGTTTAGCCCAACCAATCAGCATG 679
 Db 180 HisAsnIleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 199
 Qy 680 TCTATGTTAGTGAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGT 739
 Db 200 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 219
 Qy 740 TATAGATATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTTACTGATTCGCTTACC 799
 Db 220 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 239
 Qy 800 GCTAAACGGGTAGGTTTCAATACCTACATCGCGCGCATGTTCTTATGGTCTCTGTT 859
 Db 240 AlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 259
 Qy 860 GGTATGAAATTC 871
 Db 260 GlyTyrGluPhe 263
 RESULT 10


```
PA (CHIR-) CHIRON SPA.
XX
XX
PI Arico M, Comanducci M;
XX
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding NadA protein
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EV-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC neisserial adhesin A (NadA) allele 1/2 chimera (strain 95330)
XX
XX Sequence 355 AA;
SQ
Alignment Scores:
Pred. No.: 6.61e-16 Length: 355
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 6 Gaps: 3
US-10-030-529A-1 (1-1168) x ABU07925 (1-355)
QY 479 GATTAGACACGAGTATTTAAATTAAGTCTTGATGCACGTATTTCCAAAAATAACAA 538
DB 247 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 264
QY 539 AATATTGATCTATAAGTAAATATTTACTAGAACTGGGTACTTATTTAGATGATTCCTTAT 598
DB 265 ArgIleAspSerLeuAspLys----- 271
QY 599 CGTATGATGGACAAATAACATATATCAATAGTTGCTTAAAGATTCCAACTGGT 658
DB 272 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 283
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
DB 284 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 303
QY 719 GTTTCTGTCGGTAGGAGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
DB 304 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 323
QY 779 CGCATTACTGATCGTTTACCGCTTAACCGGGTGTAGCGTTCAATACCTACATGGC--- 835
DB 324 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 343
QY 836 GGCATGTCCTTGGGCTCTCTGTTGGTTATGAATTC 871
DB 344 SerAlaAlaTyrHisValGlyValAsnTyrGluIrrp 355
RESULT 12
ABU07926
ID ABU07926 standard; protein; 357 AA.
XX
XX AC ABU07926;
XX
XX DT 23-MAY-2003 (first entry)
XX
```

```
DE
XX
XX Neisserial adhesin A allele 1/2 chimera (first ATG) in strain 95330.
XX
XX Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;
XX neisserial infection; meningitis; bacterial meningitis; bacteraemia;
XX systemic immunity; mucosal immunity; allele; strain 95330; chimeric.
XX
XX Neisseria meningitidis.
OS Synthetic.
XX
XX WO2003010194-A2.
XX
XX 06-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-IB003396.
XX
XX 27-JUL-2001; 2001GB-00018401.
PR 06-SEP-2001; 2001GB-00021591.
PR 14-MAY-2002; 2002GB-00011025.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Arico M, Comanducci M;
XX
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding NadA protein
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EV-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC neisserial adhesin A (NadA) allele 1/2 (first ATG start) chimera (strain
CC 95330)
XX
XX Sequence 357 AA;
SQ
Alignment Scores:
Pred. No.: 6.63e-16 Length: 357
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 6 Gaps: 3
US-10-030-529A-1 (1-1168) x ABU07926 (1-357)
QY 479 GATTAGACACGAGTATTTAAATTAAGTCTTGATGCACGTATTTCCAAAAATAACAA 538
DB 249 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 266
QY 539 AATATTGATCTATAAGTAAATATTTACTAGAACTGGGTACTTATTTAGATGATTCCTTAT 598
DB 267 ArgIleAspSerLeuAspLys----- 273
QY 599 CGTATGATGGACAAATAACATATATCAATAGTTGCTTAAAGATTCCAACTGGT 658
DB 274 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 285
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
DB 286 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 305
```



```
XX PS Claim 2; Page 1377-1378; 1453pp; English.
XX PS
XX CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254773 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of the
XX CC invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polypeptides can be used in the manufacture
XX CC of medicaments for treating or preventing infection due to Neisseria
XX CC bacteria (e.g. meningitis and septicemia), to detect the presence of
XX CC Neisseria bacteria, or to raise antibodies. They may also be used to
XX CC screen for agonists or antagonists, which may themselves have use as
XX CC antibacterial agents. The polynucleotides of the invention may also be
XX CC used in gene therapy protocols
XX SQ Sequence 364 AA;

Alignment Scores:
Pred. No.: 6.68e-16 Length: 364
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: Gaps: 3

US-10-030-529A-1 (1-1168) x AA275736 (1-364)
Qy 479 GATTTAAGACACGATTTTAAATTAAGTCTTGATGCACGTATTTCACAAAATAAACA 538
Db 256 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 273
Qy 539 AATATTGATCTATAAGTAATAATTTACTAGAACTGGGTACTTATTAGATGATCTTAT 598
Db 274 ArgIleAspSerLeuAspLys----- 280
Qy 599 CGTATGATGGAACAAAATACACATAATATCAATAAGTTGTCTAAAGAAATTCGCAACATGGT 658
Db 281 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 292
Qy 659 TTAGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGTTAGGCAAAACGAGC 718
Db 293 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 312
Qy 719 GTTTCCTGCTGGTAGGAGTTATAGATATAAACTGCATTAGCATTTGGTGTGCGCTCA 778
Db 313 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 332
Qy 779 CGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAATGGC--- 835
Db 333 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 352

RESULT 15
AAU27562
ID AAU27562 standard; protein; 364 AA.
XX AC AAU27562;
XX DT 18-DEC-2001 (first entry)
XX DE Neisseria meningitidis protein 961 sequence.
XX KW Neisseria gonorrhoeae; leader peptide; fusion protein; ORF46.1;
XX KW Neisseria protein.
XX OS Neisseria meningitidis.
XX FN WO200164922-A2.

Search completed: May 13, 2004, 08:25:15
```


Job time : 69 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:18:07 ; Search time 17.5 seconds
(without alignments)
6950.622 Million cell updates/sec

Title: US-10-030-529a-1
Perfect score: 2048
Sequence: 1 ataaatcgcattgacatt.....aagccgtaaaggcgagac 1168

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2/1/USPTO_spool_p/US10030529/runat_13052004_081452_2600/app_query.fasta_1.1351
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -NATRX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529@cgn 1_16 @runat_13052004_081452_2600 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.5	9.1	455	1 YADA_YEREN	P31489 yersinia en
2	182	8.9	434	1 YADA_YEREN	P10858 yersinia ps
c 3	144	7.2	420	1 YFJD_HAETIN	Q57017 haemophilus
c 4	130	6.5	428	1 YFJD_ECOLI	P37908 escherichia
5	102.5	5.0	498	1 FLID_BACSU	P39738 bacillus su
6	101	4.9	1690	1 KFIA_HUMAN	Q12756 homo sapien
7	99.5	4.9	831	1 PPIA_SYNY3	P74397 synechocyst
8	99.5	4.9	1695	1 KFIA_MOUSE	P33173 mus musculu
9	95	4.8	795	1 LON_MYCGE	P47481 mycoplasma
10	95	4.6	2214	1 POLG_CXA24	P36290 c genome po
11	94.5	4.6	1036	1 Y946_ARCFU	Q29316 archaeoglob
12	92.5	4.5	588	1 CMC2_CASBL	Q20799 caenorhabdi
13	92.5	4.5	727	1 MEPI_ARATH	P19855 arabidopsis
14	92.5	4.5	824	1 HELI_HSV6U	P37824 human herpe
15	92	4.5	444	1 CHLB_CHLPT	P37824 chlamydomon
16	90.5	4.4	290	1 CYOA_BUCAP	Q8K993 buchnera ap
17	90.5	4.4	819	1 LON_CHLPN	Q292f4 chlamydia p
18	90.5	4.4	1226	1 DSRA_HUMAN	P55265 homo sapien

19	90.5	4.4	2716	1 OSA_DROME	Q8in94 drosophila
20	90	4.4	576	1 FTSI_BUCBP	Q89aq0 buchnera ap
21	90	4.4	758	1 PXAI_YEAST	P41909 saccharomyc
22	90	4.4	1328	1 POLX_TOBAC	P10978 nicotiana t
23	89.5	4.4	824	1 HELI_HSV6Z	P52450 human herpe
24	89.5	4.4	956	1 UVRA_LISIN	Q288a5 listeria in
25	88.5	4.3	4543	1 LRPI_CHICK	P98157 gallus gall
26	88	4.3	2206	1 POLG_POL1M	P03299 p genome po
27	87.5	4.3	643	1 TBPA_EPTST	Q90501 eptaretus
28	87.5	4.3	820	1 HELI_HSV7J	P52357 human herpe
29	87	4.2	806	1 LON_BORBU	Q59185 borrelia bu
30	87	4.2	2208	1 POLH_POL1M	P03300 p genome po
31	86.5	4.2	524	1 CLPP_CHLRE	P42380 chlamydomon
32	86	4.2	473	1 Y085_UREPA	Q3pr60 ureaplasma
33	86	4.2	677	1 YD64_MYCPN	P75417 mycoplasma
34	86	4.2	766	1 TKLI_MOUSE	Q8c0v0 mus musculu
35	86	4.2	920	1 YMLI_YEAST	Q3758 saccharomyc
36	85.5	4.2	218	1 GTH5_ARATH	P42769 arabidopsis
37	85.5	4.2	419	1 ENO_SULSO	Q97zj3 sulfobolus
38	85.5	4.2	522	1 LEUL_SHEON	Q8e9n2 shewanella
39	85	4.2	810	1 HPUB_NEIMC	P96949 neisseria m
40	85	4.2	2209	1 POLG_POL1S	P03301 p genome po
41	84.5	4.1	491	1 G6PD_BUCAI	P57405 buchnera ap
42	84.5	4.1	947	1 LKTA_PASSP	P55123 pasteurella
43	84.5	4.1	1790	1 VIT_ANTGR	Q05808 anthomomus
44	84	4.1	438	1 GLGA_THECA	P58395 thermus cal
45	84	4.1	662	1 YMEI_SCHMA	P46508 schistosoma

ALIGNMENTS

RESULT 1

YADA_YEREN	STANDARD;	PRT;	455 AA.
ID YADA_YEREN			
AC P31489;			
DT 01-JUL-1993 (Rel. 26, Created)			
DT 01-JUL-1993 (Rel. 26, Last sequence update)			
DT 01-OCT-1996 (Rel. 34, Last annotation update)			
DE Invasin precursor (Outer membrane adhesin).			
GN YADA OR YOPA OR INVA OR YOPI.			
OS Yersinia enterocolitica.			
OG Plasmid pYV.			
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC Enterobacteriaceae; Yersinia.			
OX NCBI_TaxID=630;			
RN [1]			
RC SEQUENCE FROM N.A.			
RP STRAIN=6471/76 / Serotype O:3;			
RX MEDLINE=95020586; PubMed=7934875;			
RA Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,			
RA Skurnik M.;			
RT "Hydrophobic domains affect the collagen-binding specificity and			
RT surface polymerization as well as the virulence potential of the Yada			
RT protein of Yersinia enterocolitica.";			
RL Mol. Microbiol. 10:995-1011(1993).			
CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO			
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE			
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS			
CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL			
CC SURFACE.			
CC -1- SUBCELLULAR LOCATION: Outer membrane.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC or send an email to license@sib-sib.ch).			
CC			
DR EMBL; X13882; CAA32086.1; -			
DR PIR; S04912; S04912.			
DR InterPro; IPR008126; Adhesion.			


```
ID YFJD_HAEIN STANDARD; PRT; 420 AA.
AC Q57017; P96330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0107.
GN HI0107.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO E.COLI YFJD.
CC -!- SIMILARITY: Contains 2 CBS domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32696; AAC21785.1; ALT_INIT.
DR TIGR; HI0107; -.
DR InterPro; IPR002550; CBS.
DR InterPro; IPR000644; CBS domain.
DR InterPro; IPR005170; CorC_transpt-asc.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF03471; CorC_HlyC; 1.
DR Pfam; PF01595; DUF21; 1.
DR SMART; SM00116; CBS; 1.
KW Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
SQ SEQUENCE 420 AA; 46959 MW; 7CFDF6A12937231 CRC64;

Alignment Scores:
Pred. No.: 7,41e-05 Length: 420
Score: 144.00 Matches: 29
Percent Similarity: 81.40% Conservative: 6
Best Local Similarity: 67.44% Mismatches: 8
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-030-529A-1 (1-1168) x YFJD_HAEIN (1-420)

Qy 1165 CGCACCTTTACGCTTAATTTAGACATTAGAAAAATTCGATAGAGGTACCAA 1106
Dd ArgthrPheAnsGlyLeuileLeuGluHisLeuGluLeuProAspGluGlyThrile 394
Qy 1105 TTTAGGCTAACCAACCTTAATATTACCGTTTACAGCTGGCGGATATATCGTGAAAAA 1046
Dd TTTAGGCTAACCAACCTTAATATTACCGTTTACAGCTGGCGGATATATCGTGAAAAA 1046

395 CysGluileAspGlyLeuileThrileLeuGluValGlyAspAsnMetileysGln 414
1045 GTGAAGCTA 1037
415 AlalysVal 417
|||||
RESULT 4
YFJD_ECOLI STANDARD; PRT; 428 AA.
AC P37908; P76600; P76601; P77009;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfjd.
GN YFJD OR B2612/B2613.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli K-12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE OF 322-428 FROM N.A.
RC STRAIN=Bi78;
RX MEDLINE=88319942; PubMed=3045760;
RA Lipinska B., King J., Ang D., Georgopoulos C.;
RT "Sequence analysis and transcriptional regulation of the Escherichia coli grpE gene, encoding a heat shock protein."
RL Nucleic Acids Res. 16:7545-7562(1988).
RN [4]
RP IDENTIFICATION.
RA Rudd K.E.;
RL Unpublished observations (AUG-1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO H.INFLUENZAE HI0107.
CC -!- SIMILARITY: Contains 2 CBS domains.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to numerous frameshifts.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000347; AAC75661.1; ALT_FRAME.
```

```
DR EMBL; AE000347; AAC75662.1; ALT FRAME.
DR EMBL; D90888; BAA16497.1; ALT INIT.
DR EMBL; X07863; -; NOT ANNOTATED_CDS.
DR EcoGene; EG12442; yfjD.
DR InterPro; IPR002550; CBS.
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR005170; CorC_transpt-asc.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF03471; CorC_HlyC; 1.
DR Pfam; PF01595; DUF21; 1.
DR SMART; SM00116; CBS; 1.
KW Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
SQ SEQUENCE 428 AA; 48044 MW; 38456865EDBCB151 CRC64;

Alignment Scores:
Pred. No.: 0.00124 Length: 428
Score: 130.00 Matches: 24
Percent Similarity: 81.40% Conservative: 11
Best Local Similarity: 55.81% Mismatches: 8
Query Match: 6.52% Indels: 0
DB: 1 Gaps: 0

US-10-030-529a-1 (1-1168) x YFUD_ECOLI (1-428)

Qy 1165 CGCACTTTAACGGCTTAATTTTACAACTTTAGAAAAAATTCGATAGAGTACCCAA 1106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 ArgThrValAsnGlyValIleLeuGluAlaLeuGluGluIleProValAlaGlyThrArg 393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1105 TTTAGGCTAACCACTTAATTAATACCTTTTACAGTGGCGGATATATATGTAAGAAA 1046
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 ValArgIleGlyGlyTyzAspIleAspIleLeuAspValGlnAspAsnMetIleLysGln 413
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1045 GTGAAAGTA 1037
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 414 VallysVal 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
ID FLID_BACSU STANDARD; PRT; 498 AA.
AC P39738; P96504;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Flagellar hook-associated protein 2 (HAP2) (Filament cap protein)
DE (Flagellar cap protein).
DE FLID OR MRGB OR BSU35340.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / HB2058;
RX MEDLINE=94252974; PubMed=8195064;
RA Chen L., Helmann J.D.;
RT "The Bacillus subtilis sigma D-dependent operon encoding the
flagellar proteins Flid, Flis, and Flit.";
RL J. Bacteriol. 176:3093-3101(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94252974; PubMed=8195064;
RA Soldo B., Lazarevic V., Maue C., Karamata D.;
RT "Sequence of the Bacillus subtilis 168 chromosomal region from 305
to 307 degree.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
```

```
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Avevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowaka A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis."
Nature 390:249-256(1997).
CC -1- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF
THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE
FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPPING
STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH
THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT
POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).
CC -1- SUBUNIT: Homopentamer (By similarity).
CC -1- SUBCELLULAR LOCATION: Flagellar.
CC -1- SIMILARITY: Belongs to the flid family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z31376; CAA83248.1; -.
DR EMBL; U56901; AAC44953.1; -.
DR EMBL; Z99122; CAB15551.1; -.
DR PIR; H96623; H96623.1; flid.
DR Subtilist; BG10921; flid.
DR InterPro; IPR003481; Flid.
DR Pfam; PF02465; Flid; 1.
DR Flagellum; Coiled coil; Complete proteome.
KW DONAIN 448 471 COILED COIL (POTENTIAL).
FT CONFLICT 498 AA; 54505 MW; 08254112E9564E5F CRC64;
SQ SEQUENCE 498 AA; 54505 MW; 08254112E9564E5F CRC64;

Alignment Scores:
Pred. No.: 0.319 Length: 498
Score: 102.50 Matches: 67
Percent Similarity: 33.01% Conservative: 35
Best Local Similarity: 21.68% Mismatches: 102
Query Match: 5.00% Indels: 105
DB: 1 Gaps: 14

US-10-030-529a-1 (1-1168) x FLID_BACSU (1-498)

Qy 74 GACAATTATTACTTAATGAGGTGATTATGAAATTAATGTTAGTTCGCGTAGTGGA 133
:::|||||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 242 AsnAsnPhThrValAsnGlyValThrTyrSerIleLysAsnThrThrAlaAlaThrGly 261
Qy 134 TTAGCTTGTCTACTATTACAAATGCTCAGACGCCCAAGATTGCTGGAGTATCT 193
Db 262 ProValThrThrSerValSerThr----- 269
Qy 194 TCTTTGTATAGTATGACTATGACTATGCTAAGGGTAAATGCACTTGTCTAATGAAGGC 253
Db 269 ----- 269
Qy 254 GGTTCGATATTAAGTCCAGGAGTAAATGAAGCAAGCAAGATGATTCCTAAACAG 313
Db 270 -----AspValAspGlyIleTyrAsnGlnIleLysGluPheValAspLys--- 284
Qy 314 GCTACTTATCTTGAAATTA-----CAGCATATATG 343
Db 285 -----TyrAsnGluLeuValAspSerLeuAsnGluLysLeuGluGluLysTyrArg 302
Qy 344 CCTTATCTCTGTTCTCGTACATATGCTCTGCGGTTCTCTAGCCCTATCTACTGTTA 403
Db 303 AspTyrThrProLeuThrSerGluGlnLysGluAlaMetSerAspLysGluValGluLeu 322
Qy 404 TATCGATGTCGTGCTGATCAACTCGAATAAATCGGCAGCAGCTGAATGAAATG 463
Db 323 Trp-----GluGluLysAlaLysSerGlyLeuLeuArgAsnAspSerIleSerThr 340
Qy 464 TATAGTATTTTACGATTTAAGCACGATTTTAAATTAAGTTCTTGATGCCAGTATT 523
Db 341 GlyThr-----AsnGlnMetArgThrAspPheTyrThrGlnVal----- 353
Qy 524 TCCAAATAAACAATAATGATGACTATAAGTAAATATTACTAGACTGGTACTTAT 583
Db 354 -----AsnAlaAspGlyLysThrTyrGlnLeuThrGluPheGlyIleThr 368
Qy 584 TTAGATGATTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 643
Db 369 ThrSerSerAlaTyrGlnLeuArgGlyHisLeuGluIleAsnGluLysLeuLysAla 388
Qy 644 GAATGCAAACTGGTTTATGCCAACCAATCAGCATGCTCTATGTTAGTCCAAACCAATGGT 703
Db 389 LysIleAlaGlu-----AspProGlnGly 396
Qy 704 GTAGCAAAACGAGCGTTTCTGCTGCG--GTAGGAGGTTATAGAGATAAAATGCAATTA 760
Db 397 ValAlaAsnLeuPheThrSerGlyThrAsnAspSerAsnTyrSerAspLys----- 413
Qy 761 GCCATTGTCGTCGCTCAGCATTAATGAT-----CGCTTTTACC 799
Db 414 -----GlyIleMetLysArgIleThrAsnThrLeuArgSerThrValLysSerIleGlu 431
Qy 800 GCTAAGCGGTGTAGCGTTCAATACCTACAAATGCGCGCATGCTTATGCTGCTGTT 859
Db 432 AlalysAlaGly-----AsnSerThrMetGlyAlaSerSerTyr-----Serile 446
Qy 860 GGTATGAAATCTAATCA-----TTAGTTTAA 886
Db 447 GlyLysAsnLeuAsnSerIleSerThrThrGluLeuThrAspMetGlnAspArgLeuAsnThr 466
Qy 887 ATCACTAATCGTTTGGTTATAATAAA 913
Db 467 IleGluAsnArgTyr---TyrSerLys 474
```

RESULT 6

```
KF1A_HUMAN
ID KF1A_HUMAN STANDARD; PRT; 1690 AA.
AC Q12756;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).
GN ATSV OR KIF1A
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96299637; PubMed=8661001;
RA Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;
RT "Characterization of a kinesin-related gene ATSV, within the tuberosus
sclerosis locus (TSC1) candidate region on chromosome 9q34.";
RL Genomics 33:421-429 (1996).
CC -!- FUNCTION: Motor for anterograde axonal transport of synaptic
vesicle precursors (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104
subfamily.
CC -!- SIMILARITY: Contains 1 FHA domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.ebi.ac.uk/ebis/
or send an email to license@ebi.ac.uk).
CC -----
EMBL: X90840; CAA62346.1; -.
HSSP: P11119; 3KAR.
DR Genew; HGNC:888; ATSV.
DR MIN; 601255; -.
DR GO; GO:0003774; P: motor activity; TAS.
DR GO; GO:0008089; P: anterograde axon cargo transport; TAS.
DR InterPro; IPR000253; FHA.
DR InterPro; IPR001752; Kinesin_motor.
DR InterPro; IPR001849; PH.
DR InterPro; IPR008984; SMAD_FHA.
DR Pfam; PF00498; FHA; 1.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
DR PROSITE; PS50006; FHA DOMAIN; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 361 KINESIN-MOTOR.
FT DOMAIN 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).
FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1575 1673 PH.
FT NP_BIND 97 104 ATP (POTENTIAL).
SQ SEQUENCE 1690 AA; 191083 MW; D8DDEC784624FB4D CRC64;
```

Alignment Scores:

Pred. No.:	0.447	Length:	1690
Score:	101.00	Matches:	54
Percent Similarity:	39.57%	Conservative:	39
Best Local Similarity:	22.98%	Mismatches:	92
Query Match:	4.93%	Indels:	50
DB:	1	Gaps:	11

US-10-030-529a-1 (1-1168) x KF1A_HUMAN (1-1690)

```
Qy 65 ATCAAGATGCAATATTACTTAAATGAGTGATTAATAAATTAATTTAGTTGCC 124
Db 10 ValArgValArgProPheAsnSerArgGluMetSerArgAspSerLysCysIleGln 29
Qy 125 GTATGGGATAGCTTGTCTTACTATTACAAATGGCTCAGCAGCCCAAGTTTCT 184
```


Db 167 ArgValArgGluHisProLeu---LeuglyProTyrValGluAspLeuSerLysLeuAla 185
QY 608 GAACAAATACACATAATATCAATAGTTG 637
Db 186 ValThrSerTyrAsnAspIleGlnAspLeu 195
RESULT 9
LON_MYCGE
ID LON_MYCGE STANDARD; PRT; 795 AA.
AC P47481; Q49276;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR MG239.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 484-606 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to licences@isb-sib.ch).
CC
CC EMBL; U39703; AAC71460.1; -;
CC EMBL; U02148; AAD12428.1; -;
CC PIR; D64226; D64226.
CC MEROPS; S16.004; -;
CC TIGR; MG239; -;
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.

DR SMART; SM00464; LON; 1.
DR TIGRFAMS; TIGR00763; Lon; 1.
DR PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT DOMAIN 304 308 POLY-SER.
FT NP_BIND 379 386 ATP (POTENTIAL).
FT ACT_SITE 702 702 BY SIMILARITY.
FT CONFLICT 494 494 T -> S (IN REF. 2).
SQ SEQUENCE 795 AA; 89987 MW; 31C8188EA0F5E39 CRC64;

Alignment Scores:
Pred. No.: 0.655 Length: 795
Score: 99.00 Matches: 61
Percent Similarity: 38.46% Conservative: 44
Best Local Similarity: 22.34% Mismatches: 116
Query Match: 4.83% Indels: 52
DB: 1 Gaps: 10

US-10-030-529A-1 (1-1168) x LON_MYCGE (1-795)
QY 26 AATGTAAGGTAGATAAGAAAGTAAATTCUATATTTTACAATCAAGATTGACAAATTATTTA 85
Db 163 AsnValProLysGlyLeuLysGlnLeuAspIleleThrPheLysLeuAlaAsnLeuVal 182
QY 86 CTTAATCAGGTGATTATCAAAATTTAAATGTTTGTTCGCGTAGTGGGATTAGCTTGTCT 145
Db 183 ProAsnThrGluSerIleLysGlnAlaLeuGluGlu----- 195
QY 146 ACTATTACAACATGGCTCAGCAGCCGCCAAAGTTT-----GCTGGAGTATCTCTTTG 199
Db 196 -----AsnGluIleAlaAsnArgLeuGluLysIleleGlnAlaGlyleGluAspLeu 213
QY 200 TATAGCTATGATGATGATGTAAGGTAAATGCGACTGCTCTAATGAAGCGGTTTC 259
Db 214 GlnLysleGln---AspTyrGlyArgSerLys-----AsnLysGluThrGluPhe 229
QY 260 GATATTAAAGTCCAGGATTAATAATGAAGCAAAAGAAATGGATTCTTAAACAGGCTACT 319
Db 230 AspLysLeuAspSerLysIleThrArgLysIleAsnGluGlnLeuSerArgGln----- 247
QY 320 TATCTTGAATACAGCATATATGCTTATATCTCTGTTCTCTCGGACATATGCTCTCGC 379
Db 248 -----GlnArgAspPheTyrLeuArgGluLysLeuArgIleleleArgGluGluIleGly 265
QY 380 GTTTCCTCTAGCCCTATATCTGTATATCCGATGCTGTATCTCTGATCAACTTGGATAAAT 439
Db 266 Ileser-----SerLysLysGluAspGluValalaSerIle 277
QY 440 CGGCAGCAGCTGAAATTAATTTGAT-----AGTTATTTTAAAGAT 481
Db 278 ArgLysLysLeuAspGluAsnProTyrProGluAlaIleLysLysArgIleLeuSerGlu 297
QY 482 TTAACACAGCATTTTAATAATTAAGTTCTTGATGACGCTATTTCCAAAAATAACAAAT 541
Db 298 LeuGluHisTyrGluAsnSerSerSerSerGlnGluSerThrLeuThrLysThrTyr 317
QY 542 ATTGATATCTATAAGTAAATATTTTACAGAACTG-----GGT 577
Db 318 IleaSpThr-----LeuLeuAsnLeuProTyrTrpGlnLysSerLysAspLeu 333
QY 578 ACTTATTTAGATGATTCCTTATCGTATGATGAACAAATAACACATAATATCAATAGTTG 637
Db 334 SerAspValLysAsnLeuIleLysThrLeuAspLysAsnHisThrGlyLeuAspLysVal 353
QY 638 TCTAAGAATTCGAAACTGGTTTATGCC-----AACCAATCAGCA 676
Db 354 LysGluArgIleValGluTyrLeuAlaValGlnLeuArgThrGlnLysAsnLysGlyPro 373
QY 677 TTGCTATGTTAGTGCACCAATGCTGTAGCAAAACGAGCGTTCTTCGCTCGGTAGGA 736
Db 374 IleMetCysLeuValGlyProGlyValGlyLysSerSerLeuAlaLysSerIleAla 393
QY 737 GGTATATAGAGATAAAACTGCATTAGCCATTGGTGTGCGC 775

```

Db 394 GluAlaLeuaspLysPheValLysIleSerLeugly 406
RESULT 10
POLG_CXA24
ID POLG_CXA24 STANDARD; PRT; 2214 AA.
AC P36290;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (PIA); Coat protein VP2 (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Core protein P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP3 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Coxsackievirus A24 (strain EH24/70).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=36404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271460; PubMed=1317075;
RA Supanaranond K., Takeda N., Yamazaki S.;
RT "The complete nucleotide sequence of a variant of Coxsackievirus A24, an agent causing acute hemorrhagic conjunctivitis.";
RL Virus Genes 5:149-158(1992).
CC -!- FUNCTION: It is thought that the P2C protein attaches to vesicular membranes and is associated with viral RNA synthesis.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + [RNA] (N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.
CC -!- PM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90457; -; NOT_ANNOTATED_CDS.
DR PIR; A48548; A48548.
DR HSP; P03299; 1AR7.
DR MEROPS; C03.001; -.
DR MEROPS; C03.020; -.
DR InterPro; IPR004004; Calici_pol_hel.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000199; Pept_3C_picorn.
DR InterPro; IPR000081; Peptidase_C3.
DR InterPro; IPR003138; Pico_P1A.
DR InterPro; IPR002527; Pico_P2B.
DR InterPro; IPR001676; RNV.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF02226; Pico_P1A; 1.
DR Pfam; PF00947; Pico_P2A; 1.
DR Pfam; PF01552; Pico_P2B; 1.
DR Pfam; PF00073; rhv_3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
PRINTS: PR00918; CALICVIRUSNS.
ProDom; PD001125; Cys_protease_3C; 1.
ProDom; PD001306; Pico_P2A; 1.
ProDom; PD001274; Pico_P2B; 1.
KW Polyprotein; Coat protein; Core protein; Transferase; RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate; Lipoprotein.
KW CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 340 COAT PROTEIN VP2.
FT CHAIN 341 577 COAT PROTEIN VP3.
FT CHAIN 578 888 COAT PROTEIN VP1.
FT CHAIN 889 1035 CORE PROTEIN P2A.
FT CHAIN 1036 1132 CORE PROTEIN P2B.
FT CHAIN 1133 1461 CORE PROTEIN P2C.
FT CHAIN 1462 1548 CORE PROTEIN P3A.
FT CHAIN 1549 1570 GENOME-LINKED PROTEIN VP3.
FT CHAIN 1571 1753 PICORNAIN 3C.
FT CHAIN 1754 2214 RNA-DIRECTED RNA POLYMERASE.
FT ACT_SITE 1717 1717 N-myristoyl glycine (in host) (By similarity).
FT ACT_SITE 1731 1731 PROTEASE (POTENTIAL).
SQ SEQUENCE 2214 AA; 247212 MW; E0DD74569E1B22B8 CRC64;
Alignment Scores:
Pred. No.: 1.51 Length: 2214
Score: 95.00 Matches: 70
Percent Similarity: 35.05% Conservative: 46
Best Local Similarity: 21.15% Mismatches: 110
Query Match: 4.64% Indels: 105
DB: 1 Gaps: 17
US-10-030-529A-1 (1-1168) x POLG_CXA24 (1-2214)
QY 127 AGTGGGATTAGTGTCTTCTACTATTACACAATGGCTCAGCAGCGCAAG----- 178
DB 761 SerAsnProSerValPheThyThrTyr-----GlySerAla-ProProArgIleSerIl 778
QY 179 -----TTTGTCTGGAGTATCTTCTTTGTATAGTATGATGATGAC---TATGTAAG----- 226
DB 778 eProTyrValGlyIleAlaAlaSerHisPheThyAspGlyPheAlaArgValPr 798
QY 227 -----GGTAAATGGTCTTGCTTAA 246
DB 798 oLeuLysAspGluThrValAspSerGlyAspThrTyrGlyLeuValThrIleAsnAs 818
QY 247 TGAAGGCGGTTTCGATATTAAAGTCCAGCGGATTAAATGAAGCCAAAGAAATGGATTTC 306
DB 818 pPheGlyThrLeuAlaValargValValAsn---GluPheAsnProAlaargIleIleSe 837
QY 307 TAAACAGGCTACTTATCTTGAATTACAGCAT-----TATATGCTTATATCTCTGT 357
DB 837 xLysIleArgValTyrMetLysProLysHisValArgCysTrpCysProArgProAr 857
QY 358 TCTCGTGACATAT---GCTCTCGCGTTTCTCTAGCCCTATCTACTGTATATCGATGTC 414
DB 857 gAlaValProTyrArgGlyGluGlyValAspPheLysGlnAspSerIleThrProLeuIl 877
QY 415 TGATCTCTGATCAACTT-----GGAATAAATCGGCAGCAGCTG----- 451
DB 877 eAlaValGluAsnIleAsnThrPheGlyGlyPheGlyHisGlnAsnMetAlaValTyrVa 897
QY 452 -----AAATTGATTTGTATAGTATTTTAAAGATTTTAAAGACACACATTTTAAATT 501
DB 897 lAlaGlyTyrLysIleCysAsnTyrHisLeuAlaThrProGluAspHisAspAsnAlaVa 917
QY 502 AAAAGTCTTGATGACGT-----ATTCCAAAAATAAACAATAATTATTGATAC 549
DB 917 lArgValLeuTrpAsnArgAspLeuMetIleValSerSerArgAlaGlnGlySerAspTh 937
QY 550 TATAAGTAAATATTACTAGAACCTGGTACTTAT----- 583
DB 937 rIleAlaArgCysAsnCysArgThrGlyValTyrCysLysSerMetLysLysTyrTy 957

```


XX PS Claim 7; Page 58-59; 80pp; English.

XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus

XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance

XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting

XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA

XX CC polypeptide is used to produce vaccine compositions, which are useful for

XX CC inducing a protective immune response in a subject at risk of developing

XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic

XX CC fragments or peptides can be used for screening libraries of compounds in

XX CC a variety of drug screening techniques. The proteins and peptides may be

XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in

XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc..

XX CC The nucleic acids are useful for the preparation of DsrA proteins

XX CC

SQ Sequence 257 AA;

Alignment Scores:

Pred. No.: 7.02e-148 Length: 257

Score: 1342.00 Matches: 257

Percent Similarity: 100.00% Conservativity: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 65.53% Indels: 0

DB: 4 Gaps: 0

US-10-030-529A-1 (1-1168) x AAB31713 (1-257)

QY 101 ATGAAATTAATGTTAGTTGGCTAGTGGATTAGCTTCTTCTACTATTACAAATG 160

DB 1 MetLysileYsLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20

QY 161 GCTCAGCAGCGCCAAAGATTTGCTGGAGTATCTTTTGTATAGCTATGACTAT 220

DB 21 AlaGlnProProLysPheAlaGlyValSerLeuTyrSerTyrGluTyrAspTyr 40

QY 221 GGTAAAGGTAATCGACTGTCTTAATGAAGCGGTTTCGATATTAAGTCCAGGATT 280

DB 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyPheAspIleLysValProGlyIle 60

QY 281 AAAATGAAGCAAAAGATGATTTCTAAACAGCGTACTTATCTTGAATTACACATTAT 340

DB 61 LysMetLysProLysGluTrpLysSerLysGlnAlaThrLysLeuGluLeuGlnHisTyr 80

QY 341 ATGCCTTATCTCTGTTCTCTGCACATATGCTCTCGCGTTTCTCTAGCCCTATATG 400

DB 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100

QY 401 TTATATCGATGCTGATCCTGATCACTTGGATAAATCGGACGAGCTGAAATTTGAAT 460

DB 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120

QY 461 TTGTATAGTATTTTAAAGATTAAAGCACCATTTTTAAATTTAAAGTTCTTGTATGCGGT 520

DB 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140

QY 521 ATTTCCAAAAATAACAAAATATTGATCTATAAGTAAATATTACTAGAACTCGGTACT 580

DB 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160

QY 581 TATTAGATGATTTCTTATCGTATAGTGAACAAATACATATATATCAATTAAGTTGTCT 640

DB 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180

QY 641 AAGAATTCGAACCTGTTTACCAACCAATCAGCATTTGTATGTAGTCCACCAAT 700

DB 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200

QY 701 GGTGTAGCAAAACAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAAATGCATTA 760

DB 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220

QY 761 GCCATTGGTCTCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTC 820

DB 221 AlaileGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240

QY 821 AATACCTACAAATCGCGCATGTCTTATGTGCTGCTTCTTGTATGAATTC 871

DB 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257

RESULT 2

AAB31705

ID AAB31705 standard; protein; 257 AA.

XX AAB31705;

AC AAB31705;

XX 30-APR-2001 (first entry)

DT

DE Amino acid sequence of the DsrA locus.

XX

XX DsrA protein; outer membrane protein; serum resistance; vaccine;

XX chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX N-PSDB; AAF25262.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum

XX resistance to the bacteria used to produce vaccines that induce immune

XX response against the bacteria in subject at risk of developing chancroid.

XX Claim 3; Fig 3; 80pp; English.

XX

XX The present sequence represents a DsrA 30 kDa protein of Haemophilus

XX ducreyi. DsrA is an outer membrane protein that confers serum resistance

XX to the bacteria. DsrA antisense oligonucleotides are useful for detecting

XX a polynucleotide which encodes DsrA in a biological sample. The DsrA

XX polypeptide is used to produce vaccine compositions, which are useful for

XX inducing a protective immune response in a subject at risk of developing

XX Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

XX transmitted by sexual contact. DsrA, its catalytic or immunogenic

XX fragments or peptides can be used for screening libraries of compounds in

XX a variety of drug screening techniques. The proteins and peptides may be

XX used as antigens in immunoassays for detection of Haemophilus ducreyi in

XX various tissues and body fluids, e.g. blood, spinal fluid, sputum etc..

XX The nucleic acids are useful for the preparation of DsrA proteins

XX

SQ Sequence 257 AA;

Alignment Scores:

Pred. No.: 7.02e-148 Length: 257

Score: 1342.00 Matches: 257

Percent Similarity: 100.00% Conservativity: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 65.53% Indels: 0

DB: 4 Gaps: 0

US-10-030-529A-1 (1-1168) x AAB31705 (1-257)

QY 101 ATGAAATTAATGTTAGTTGGCTAGTGGATTAGCTTCTTCTACTATTACAAATG 160

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:17:32 ; Search time 64 Seconds
(without alignments)
10312.984 Million cell updates/sec

Title: US-10-030-529a-1

Perfect score: 2048

Sequence: 1 ataatacgtcattgacatt.....aagccgttaaaggtcgagc 1168

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xl
-Q=/cgn21/USFTO_spool_p/US10030529/runat_13052004_081451_2590/app_query.fasta_1.1351
-DB=A_Geneseq_29Jan04 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529 @CGN 1.1 81 @runat_13052004_081451_2590 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1342	65.5	257	4 AAB31713	Aab31713 Amino aci
2	1342	65.5	257	4 AAB31705	Aab31705 Amino aci
3	1342	65.5	257	4 AAB31709	Aab31709 Amino aci
4	1299.5	63.5	264	4 AAB31706	Aab31706 Amino aci
5	1299.5	63.5	264	4 AAB31707	Aab31707 Amino aci
6	1289	62.9	271	4 AAB31711	Aab31711 Amino aci
7	1283	62.6	273	4 AAB31712	Aab31712 Amino aci
8	1259	61.5	242	4 AAB31708	Aab31708 Amino aci
9	1205	58.8	263	4 AAB31710	Aab31710 Amino aci
10	218	10.6	323	6 ABU07927	Abu07927 Neisseria

11	217	10.6	355	6 ABU07925	Abu07925 Neisseria
12	217	10.6	357	6 ABU07926	Abu07926 Neisseria
13	217	10.6	362	6 ABU07915	Abu07915 Neisseria
14	217	10.6	364	3 AAY75736	Aay75736 Neisseria
15	217	10.6	364	4 AAU27562	Aau27562 Neisseria
16	217	10.6	364	6 ABU07918	Abu07918 Neisseria
17	214	10.4	325	6 ABU07928	Abu07928 Neisseria
18	209	10.2	391	6 ABU07921	Abu07921 Neisseria
19	209	10.2	393	6 ABU07922	Abu07922 Neisseria
20	209	10.2	398	6 ABU07916	Abu07916 Neisseria
21	209	10.2	400	6 ABU07919	Abu07919 Neisseria
22	204	10.0	405	6 ABU07917	Abu07917 Neisseria
23	204	10.0	405	6 ABU07923	Abu07923 Neisseria
24	204	10.0	407	6 ABU07924	Abu07924 Neisseria
25	204	10.0	407	6 ABU07920	Abu07920 Neisseria
26	204	10.0	645	4 AAE10036	Aae10036 N. mening
27	204	10.0	645	4 AAU27601	Aau27601 Neisseria
28	204	10.0	648	4 AAE10028	Aae10028 N. mening
29	204	10.0	648	4 AAU27579	Aau27579 Neisseria
30	204	10.0	793	4 AAE10016	Aae10016 N. mening
31	204	10.0	793	4 AAU27571	Aau27571 Neisseria
32	204	10.0	806	4 AAE10033	Aae10033 N. mening
33	204	10.0	806	4 AAE10035	Aae10035 N. mening
34	204	10.0	806	4 AAU27598	Aau27598 Neisseria
35	204	10.0	806	4 AAU27600	Aau27600 Neisseria
36	204	10.0	858	4 AAE10019	Aae10019 N. mening
37	204	10.0	858	4 AAU27574	Aau27574 Neisseria
38	204	10.0	1444	4 AAE10037	Aae10037 N. mening
39	204	10.0	1444	4 AAU27602	Aau27602 Neisseria
40	204	10.0	1447	4 AAE10024	Aae10024 N. mening
41	204	10.0	1447	4 AAU27577	Aau27577 Neisseria
42	193	9.4	610	2 AAU68206	Aaw68206 M. catarr
43	193	9.4	624	2 AAU68204	Aaw68204 M. catarr
44	193	9.4	889	2 AAU68208	Aaw68208 M. catarr
45	190.5	9.3	573	2 AAU68202	Aaw68202 M. catarr

ALIGNMENTS

RESULT 1
AAB31713
ID AAB31713 standard; protein; 257 AA.
XX
AC AAB31713;
DT 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain OF406.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UYN-C-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI: 2001-138311/14.
DR N-PSDB; AAF25270.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.

Db 1 MetLysIleLysCysLeuValAlaValAlaGlyLeuAlaCysSerThrIleThrThrMet 20
 QY 161 GCTCAGACGCGCCAAAGTTGCTGGAGTATCTCTTGTATAGCTATGATGACTAT 220
 Db 21 AlaGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
 QY 221 GGTAAAGGTAATGCACTTGTCTAAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
 Db 41 GlyLysGlyLysThrThrTrpSerAsnGluGlyPheAspIleLysValProGlyIle 60
 QY 281 AAATGAAGCCAAAGAAAGTATCTTAACAGGCTTATCTTGAATACACATAT 340
 Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
 QY 341 ATGCTTATATCTCTGCTCGTACATATGCTCTGCGCTTCTCTAGCCCTATCTG 400
 Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
 QY 401 TTATATCCGATGCTGATCTGATCAACTTGAATAAATCGGACGAGCTGAATTCGAA 460
 Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
 QY 461 TTGTATAGTTATTTAAACGATTTAAGACACGATTTTAAATGAAAGTTCTTGATG 520
 Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 QY 521 ATTTCCAAAATAACAAAATATTCATATAGTAAATATTTACTAGAACTGGGTACT 580
 Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
 QY 581 TATTAGATGATCTTATCGTATCATGGAACAAATACACATAATATCAATAGTCTGTCT 640
 Db 161 TyrLeuAspAspSerTyrArgMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
 QY 641 AAGAATTGCAACTGGTTAGCCCAACCAATTCAGCATTTGTCTATGTTAGTGCAACCAAT 700
 Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
 QY 701 GGTGTAGCAAAACGACGCTTCTGCTGCGTAGGAGGTATAGAGATAAATGCAATTA 760
 Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
 QY 761 GCCATTGCTGCGCTCACGATTCATGATCGCTTTACCGCTTAAGCGGGTGTAGCGTTC 820
 Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
 QY 821 AATACCTACAAATGCGCGCATGCTTATGCTGCTTCTGTTGTTATGAATTC 871
 Db 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257
 RESULT 3
 ID AAB31709
 AC AAB31709
 AC AAB31709;
 XX 30-APR-2001 (first entry)
 DT Amino acid sequence of the DsrA protein from strain CIP542 (CDC).
 DE DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 XX Haemophilus ducreyi.
 OS WO200104138-A1.
 PN 18-JAN-2001.
 PD 07-JUL-2000; 2000WO-US018834.
 XX 09-JUL-1999; 99US-0143257P.
 PR 581 TATTAGATGATCTTATCGTATGATGGAACAAATACACATAATATCAATAGTCTGTCT 640

(UNNC-) UNIV NORTH CAROLINA.
 (ELKI/) ELKINS C.

Elkins C;

WPI; 2001-138311/14.

N-PSDB; AAF25266.

Novel purified DsrA protein of Haemophilus ducreyi which confers serum resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.
 Claim 7; Page 55; 80pp; English.

The present sequence represents a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of Haemophilus ducreyi in various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. . The nucleic acids are useful for the preparation of DsrA proteins

SQ Sequence 257 AA;

Alignment Scores:
 Pred. No.: 7,02e-148 Length: 257
 Score: 1342.00 Matches: 257
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 65.53% Indels: 0
 DB: 4 Gaps: 0

US-10-030-529A-1 (1-1168) x AAB31709 (1-257)

QY 101 ATGAAATTAATTAATGTTAGTTGCGTAGTGGGATAGCTTGTCTACTATTACAAATG 160
 Db 1 MetLysIleLysCysLeuValAlaValAlaGlyLeuAlaCysSerThrIleThrThrMet 20
 QY 161 GCTCAGACGCGCCAAAGTTGCTGGAGTATCTCTTGTATAGCTATGATGACTAT 220
 Db 21 AlaGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
 QY 221 GGTAAAGGTAATGCACTTGTCTAAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
 Db 41 GlyLysGlyLysThrThrTrpSerAsnGluGlyPheAspIleLysValProGlyIle 60
 QY 281 AAATGAAGCCAAAGAAATGCACTTCTAAACAGGCTTATCTTGAATACACATAT 340
 Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
 QY 341 ATGCTTATATCTCTGCTCGTACATATGCTCTGCGCTTCTCTAGCCCTATCTG 400
 Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
 QY 401 TTATATCCGATGCTGATCTGATCAACTTGAATAAATCGGACGAGCTGAATTCGAA 460
 Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
 QY 461 TTGTATAGTTATTTAAACGATTTAAGACACGATTTTAAATGAAAGTTCTTGATG 520
 Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 QY 521 ATTTCCAAAATAACAAAATATTCATATAGTAAATATTTACTAGAACTGGGTACT 580
 Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
 QY 581 TATTAGATGATCTTATCGTATGATGGAACAAATACACATAATATCAATAGTCTGTCT 640


```
|||||
161 TyrLeuAspSerTyrArgMetMetGluGlnAsnThrHisAsnLysLeuSer 180
|||||
641 AAAGAATTGCAACTGGTTTACCAACCAATCAGCATGTCTATGTAGTGCACCAAT 700
|||||
181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
|||||
701 GGTCTAGGCAAAACGAGGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATTA 760
|||||
201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
|||||
761 GCCATTGGTGTGCGCTCACGCAATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTC 820
|||||
221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
|||||
821 AATACCTACAAATGGCGCATGCTCTTATGGTGCTCTCTGTGGTTATGAATTC 871
|||||
241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257
|||||
RESULT 4
AAB31706
ID AAB31706 standard; protein; 264 AA.
XX AC AAB31706;
XX DT 30-APR-2001 (first entry)
XX DE Amino acid sequence of the DsrA protein from strain CIPA75.
XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX KW chancroid disease; genital ulcer disease.
XX OS Haemophilus ducreyi.
XX FN WO200104138-A1.
XX PD 18-JAN-2001.
XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX PA (UYNC-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX FI Elkins C;
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25263.
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX PT resistance to the bacteria used to produce vaccines that induce immune
XX PT response against the bacteria in subject at risk of developing chancroid.
XX PS Claim 7; Page 53; 80pp; English.
XX CC The present sequence represents a DsrA 10 kDa protein of Haemophilus
XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX CC polypeptide is used to produce vaccine compositions, which are useful for
XX CC inducing a protective immune response in a subject at risk of developing
XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX CC fragments or peptides can be used for screening libraries of compounds in
XX CC a variety of drug screening techniques. The proteins and peptides may be
XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
XX CC The nucleic acids are useful for the preparation of DsrA proteins
XX SQ Sequence 264 AA;
Alignment Scores:
```

```
Pred. No.: 6.83e-143 Length: 264
Score: 1299.50 Matches: 253
Percent Similarity: 95.83% Conservative: 0
Best Local Similarity: 95.83% Mismatches: 4
Query Match: 63.45% Indels: 7
DB: 4 Gaps: 1
US-10-030-529A-1 (1-1168) x AAB31706 (1-264)
Qy 101 ATGAAATTAATAGTTTAGTTGCGGTAGTGGATTAGCTTGTTCTACTATTACACAAATG 160
Db 1 MethLysIleLysCysLeuValAlaValValGlyLeuAlaCysSerThrIleThrMet 20
Qy 161 GCTCAGCAGCGCCAAAGTTTGTGGAGTATCTTCTTTGTATAGCTATGAGTAGTACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Qy 221 GGTAAGGGTAAATCGACTTCGTCTAATGAAGGGGGTTTCGATATATAAGTCCAGGGATT 280
Db 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Qy 281 AAAATGAAGCCAAAGAAATCGATTCTTAAACAGGCTACTTATCTTGAATTACAGCATTAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Qy 341 ATGCTTATATCTCTGTTCTCGTGACATATGCTCTCGCGGTTTCTCTAGGCCCTATACTG 400
Db 81 MetProTyrThrProValLeuValThrTyrAlaHisAspValProProSerSerIleLeu 100
Qy 401 TTATATCCGATGTCTGATCCTGATCAACTTGGATAAATCGGACGACGCTGAATTAATGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
Qy 461 TTGTATAGTTATTTTAACGATTTAAGACACAGATTTTAAATTAAGTCTTCTGTATGCAGGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
Qy 521 ATTTCCAAAAATAACAAAAATATTGATACATATAGTAAATATTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
Qy 581 TATTAGATGATTTCTTATCGTATGATCGAAACAA-----AATACA 619
Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180
Qy 620 CATATATCAATAAGTTGTCTTAAAGAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTG 679
Db 181 HisAsnIleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 200
Qy 680 TCTATGTTAGTCAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGT 739
Db 201 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 220
Qy 740 TATAGAGATAAACTGCATTAGCCATTGGTGTGGCTCACGCATTACTGATCGCTTTACC 799
Db 221 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 240
Qy 800 GCTTAAACGGGTGTAGGTTCAATACCTACATCGCGGCATGTCTTATGGTCTCTGTT 859
Db 241 AlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 260
Qy 860 GGTTCATCAATTC 871
Db 261 GlyTyrGluPhe 264
RESULT 5
AAB31707
ID AAB31707 standard; protein; 264 AA.
XX AC AAB31707;
XX DT 30-APR-2001 (first entry)
XX
```


[illegible]

```

RESULT 12
CMC2_CAEEL          STANDARD;          PRT;      588 AA.
ID   CMC2_CAEEL
AC   Q20759;
DC   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable calcium-binding mitochondrial carrier F55A11.4.
GN   F55A11.4.
OS   Caenorhabditis elegans.
OC   Sukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RS   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Kershaw J.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: Calcium-dependent mitochondrial solute carrier (By
CC   similarity).
CC   -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC   inner membrane (By similarity).
CC   -1- SIMILARITY: Belongs to the mitochondrial carrier family.
CC   -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC   -1- SIMILARITY: Contains 3 Solcar repeats.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC   EMBL; Z72511; CAA96658.3; --
CC   PIR; T22688; T22688.
DR   WormPep; F55A11.4; CE05946.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR002067; Mit carrier.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF00133; mito_carr; 3.
DR   PRINTS; PR00926; MITOCARRIER.
DR   PRODOM; PD000012; EF-hand; 1.
DR   SMART; SM00054; EFh; 3.
DR   PROSITE; PS00018; EF_HAND; 2.
DR   PROSITE; PSS0920; SOLCAR; 3.
KW   Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
KW   Transmembrane; Transport; Calcium-binding.
FT   TRANSGEM 252 269
FT   POTENTIAL 1

```

FT	TRANSMEM	307	326	2 (POTENTIAL) .
FT	TRANSMEM	352	365	3 (POTENTIAL) .
FT	TRANSMEM	403	422	4 (POTENTIAL) .
FT	TRANSMEM	446	463	5 (POTENTIAL) .
FT	TRANSMEM	504	523	6 (POTENTIAL) .
FT	CA_BIND	52	63	EF-HAND 1.
FT	CA_BIND	86	97	EF-HAND 2.
FT	DOMAIN	123	133	ANCESTRAL CALCIUM SITE 3.
FT	CA_BIND	153	164	EF-HAND 4.
FT	REPEAT	246	332	SOLCAR 1.
FT	REPEAT	342	428	SOLCAR 2.
FT	REPEAT	440	529	SOLCAR 3.
SQ	SEQUENCE	588 AA;	66336 MW;	68DDF60923D8697D CRC64;

Alignment Scores:	
Pred. No.:	2.4
Score:	92.50
Percent Similarity:	41.59%
Best Local Similarity:	22.43%
Query Match:	4.52%
DB:	1

US-10-030-529A-1 (1-1168) x CMC2_CABEL (1-588)	
Qy 245 AATGAAGCGGTTTCGATATTAAGGTCAGGGATTAAATGAAGCCAAAGATGGATT	304
Db 89 AsnAspGlyThrIleAspIleArg	96
Qy 305 TCTAAACAGGCTACTTATCTTGAATTACAGCAATTATATGCCTTATACCTGTCTCTGCTCGTG	364
Db 97 -----AspLeuThrLeuAlaLeuAlaGlyHisGluThrProHisIlePro	110
Qy 365 ACATATGCTCTGGCGTTTCTCTAGCCCTACTACTGTATATCCGATCTGTATCGATCTGAT	424
Db 111 -----AlaAsnLeuAlaProValIleMetSerLysMetSer	124
Qy 425 CAACCTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAAC	481
Db 125 AspGluGly	138
Qy 482 TTAAGACACGAGTATTTAAATTAAGTTCTCTGATGCACGATATTTCCAAAAATAAACAAAT	541
Db 139 LeuGluAsnGluGlnLysLeuAlaGluMetPheAlaAspMetAspArgAsnHisAspGly	158
Qy 542 ---ATTGATACTATA-----AGTAATAATTTTACTAGAACTCGGTACTATTAGATGAT	592
Db 159 LeuValAspValValGluMetLysAsnTyrCysLysAspIleGlyValProLeuAspAsp	178
Qy 593 TCTTATCGTATGATGGAAACAAATACACATAATATCAATAAGTTGTCTAAAGAATTGCCAA	652
Db 179 -----HisLysAlaGlnHisIleValAsnLysMetAsp	190
Qy 653 ACTGGTTTATGCGCAACCAATCAGCATTTGCT-----ATGTTAGTGCACCAAAAT	700
Db 191 ThrGlySerAlaSerValAspLeuLysGluPheGlnGluPheMetMetLeuTyrProSer	210
Qy 701 GGTGTAGGCAAAACGAGCGTTCTCTCGCGTAGGAGGTTATAGAGATAAAACTGCAITTA	760
Db 211 -----SerAspLeuLysAspIleValAspPheTrpArgHisAsnLeuIleIle	226
Qy 761 GCCATTGGTGTGGCTCAGCGATTTACTGATCGCTTTACC-----GCTAAAGCGGGT	811
Db 227 AspIleGlyGluAspSerGlnIleProGluAspPheSerGlnGlnGluMetGlnGluGly	246
Qy 812 GTAGCGTTCAATACCTCAATATGGCGGCATGCTTATGTGCT	853
Db 247 IletIrpIrpArgHisLeuValAlaGlyGlyValAlaGlyValAla	260

Db	157	GluThrIleGluSerLeuLysAsnGlnLeuLysAspArgGluArgAlaLeuValLeuLys	176
Qy	461	TTGTATAGTATTTTAAACGATTTTAAAGACACGATTTTAAAGTCTTCTGATGCACGT	520
Db	177	GluLysAspPheGluAlaLysLeuGlnHisGluGlnGluGluGluGluValGlu	196
Qy	521	ATTTCCTCAAAATAACAA	565
Db	197	LysAlaLysGluGlnSerLeuIleAsnGlnLeuAsnSerAlaLysAspLeuVal	216
Qy	566	CTAGAACTGGGTACTTATTAGATGATTCCTATCGTATGATGGACAAATAACACATAAT	625
Db	217	ThrGluLeuGlyArgGluLeuSerGluLysLysLeuCysGluLysLeuLysAspGln	236
Qy	626	ATCAATAAGTGTCTCTAAAGATTGCAAACTGGTGTAGCCAAACCAATCAGCATTTGCTATG	685
Db	237	IleGluSerLeuGluAsnSerLeuSerLysAlaGlyGluAspLysGluAlaLeuGluThr	256
Qy	686	TTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGTCGGTAGGAGTTATAGA	745
Db	257	LysLeu	259
Qy	746	GATAAACTGCATTACCATTTGGTGTGCGGCTCAGCATTT	784
Db	260	GluLysLeuAspLeuValGluGlyLeuGlnAspArgIle	272
RESULT 14			
Db	HELI_HSV6U	STANDARD;	PRT; 824 AA.
AC	P52356;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Probable helicase.		
OS	U77 OR HRF2.		
GN	Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Betaherpesvirinae; Roseoloviruses.		
OX	NCBI_TaxID=10370;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=95266321; PubMed=7747482;		
RA	Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,		
RA	Martin M.E., Estathou S., Craxton M., Macaulay H.A.;		
RT	"The DNA sequence of human herpesvirus-6: structure, coding content,		
RT	and genome evolution."		
RL	Virology 209:29-51(1995).		
CC	-1- FUNCTION: THIS PROTEIN MAY BE AN HELICASE AND IS REQUIRED FOR		
CC	REPLICATION OF VIRAL DNA.		
CC	-1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL5,		
CC	EHV-1 57, EBV BBLP4, HCMV UL105, AND VZV 55.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U13194; AAA68468.1; -		
DR	EMBL; X83413; CAA58369.1; -		
DR	InterPro; IPR003840; Herpes_helicase.		
DR	Pfam; PF02689; Herpes_Helicase; 1.		

KW DNA replication; ATP-binding; Helicase.
FT NP_BIND 90 97 ATP (BY SIMILARITY)
SQ SEQUENCE 824 AA; 93286 MW; 5A5057544E06184C CRC64;

Alignment Scores:

Pred. No.: 2.43 Length: 824
Score: 92.50 Matches: 53
Percent Similarity: 38.75% Conservative: 40
Best Local Similarity: 22.08% Mismatches: 94
Query Match: 4.52% Indels: 53
DB: 1 Gaps: 12

US-10-030-529A-1 (1-1168) x HELI_HSV6U (1-824)

```
QY 329 TTACAGATTAT-----ATGCTTATATCTCTGTCGTGACATATGCTCCTGGC 379
D 315 LeuLysHisIleGluPheGlyLeuProLeuLysProGluLeuIleGluTyrValAspGln 334
QY 380 ---GTTTCTCTAGCCCTATATCTGTTATCCGATGCTGATCTGAT-----CAACTT 430
D 335 PheValLysProAlaSerTyrIleArgAsnPromMetAsnGluIleGluThrArgLeu 354
QY 431 GGAATAATCGGACGCTGAAATGTTATAGTTATTTTAAAGATTAAAGACAC 490
D 355 PheLeuSerHisAsnGluValLys-----AsnTyrPheArgSerLeuHisGlu 370
QY 491 GATTTTAAATTAAGTCTTGATGACGATTTTCCAAAATAAACAATAATT----- 544
D 371 GlnValGluVal-----ThrAsnArgAsnAsnLeuPheVal 382
QY 545 -----GATACTATAAGTAAATATTTACTAGAACTGGGT 577
D 383 PheProValTyrPheLeuIleLysAsnLysThrPheGluAspTyrLysSerGluIleGly 402
QY 578 ACTTATTAGATGATTTCTATCGTATGATGACAAACATACATATATCAATAAGTTG 637
D 403 AsnPheSerLeuGluIleGluProTyrPheLysSerAsnIleHisArgLeuAsnThrTyr 422
QY 638 TCTAAGAAATGCAAACTGGTTAGCCACCAATCAGCATGCTGTCTAGTTAGTGCACCA 697
D 423 SerGlnPheAlaAspGlnAspLeuSerLysThrValGlnLeuGluGluIleValLeuGlu 442
QY 698 AATGGT---GTAGCAAAACGAGGTTTCTGCTCGGTAGGAGGTATAGAGATAAAACT 754
D 443 AspGlySerValGluGluThrLeuIleThrCysHisLeuLysHisIleArgAsnSer--- 461
QY 755 GCATTAGCATGTTGTCGGCTCAGCATTAATGCTGCTTACCGCTTAAAGCGGCTGTA 814
D 462 -----SerIleGlyValThrSerLysIle-----LysAlaSer--- 472
QY 815 CGGTTCAATACCTACAATGCGCGCATGCTTATGCTGCTTCTGTTGTTATGAA---TTC 871
D 473 -----ThrValGlyPheSerGlyThrTyrGluLysPhe 483
QY 872 TAATCATTA-----CGTTTAACTACATCGTTTGGTTTATATAAAAGGCTAAA 922
D 484 ValGluLeuLeuGlnSerAspPheIleGluLysThrSerCysAspGlnThrIleHis 503
QY 923 TGTTTCTCTCCATCATATTAGCCCTTTCTTATTATCTTTGTTATAGCTTTGCTGTTATAA 982
D 504 AlaTyrSerPheLeuSerGlyLeuMetPheGlyGlyMetTyrSerPheCysSerLys 523
```

RESULT 15

CHLB_CHLPT STANDARD; PRT; 444 AA.
AC P37824;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit B
DE (EC 1.18.-.-) (LI-POR subunit B) (DPOR subunit B) (Fragment).
GN CHLB.
OS Chlamydomonas pilschmannii.

Chloroplast.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3061;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9403309; PubMed=8219066;
RA Liu X.-Q., Xu H., Huang C.;
RT "Chloroplast chlB gene is required for light-independent chlorophyll
accumulation in Chlamydomonas reinhardtii.";
RL Plant Mol. Biol. 23:297-308(1993).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
protochlorophyllide (pchlide) to form chlorophyllide a (Chlide)
CC (by similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
of three subunits; chlB, chlN and chlB. Could form a
heterotrimer of two chlB and two chlN subunits (By similarity).
CC -!- SIMILARITY: Belongs to the chlB / bchB / bchZ family.
DR HAMAP: MF 00353; atypical; 1
DR InterPro: IPR000510; Oxidized_nitro; 1.
DR Pfam: PF00148; oxidized_nitro; 1.
KW Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
FT NON_TER 444 444
SQ SEQUENCE 444 AA; 50058 MW; 6906E52CAAE4857F CRC64;

Alignment Scores:
Pred. No.: 2.64 Length: 444
Score: 92.00 Matches: 60
Percent Similarity: 39.23% Conservative: 42
Best Local Similarity: 23.08% Mismatches: 98
Query Match: 4.49% Indels: 61
DB: 1 Gaps: 13

US-10-030-529A-1 (1-1168) x CHLB_CHLPT (1-444)
QY 38 AATAAGAAAGTAATTCATATTTACAATCAAGATTGACAAATTTATTTAATGAGGTG 97
D 124 AspLysAlaThrSerAspValIleLeuAlaAspValAsnHisTyrArgValAsnGluLeu 143
QY 98 ATTATGAAAATTAATGTTTAGTTGCGGTAGT-----GGA 133
D 144 GlnAlaAlaAspArgThrLeuGluGlnIleValArgPheTyrLeuGluLysGluLys 163
QY 134 TTAGCTGTTCTACTATTACAATGCTGACAGCGC---CCAAAGTTTGTGAGTA 190
D 164 LeuAsnThrAsnThrIleProThrLysThrLysProSerAlaAsnIleIleGlyLeu 183
QY 191 TCTCTTTG---TATAGCTATGATGATGTTAAGGTAAATGACTTGTCTTAAT 247
D 184 PheThrLeuGlyPheHisAsnGlnHisAspCysArgGluLeuLysArgLeuLeuAsnAsn 203
QY 248 -----GAAGCGGTTTCGATATATAAAGTCCAGGG 277
D 204 LeuGlyIleGluValAsnGluIleIleProGluGlyGly-----SerValThrAsn 220
QY 278 ATTAAGTAAGCCAAAGAAATGATTTCT-----AAA 310
D 221 LeuLysAsnLeuProHisAlaTyrPheAsnLeuValProTyrArgGluIleGlyLeuMet 240
QY 311 CAGCTACTTATCTTGAATTTACAGCATATATGCTTATCTCTGTTCTGTCACATAT 370
D 241 ArgAlaValTyrLeuGluLysGluPheAsnMetProTyr-----ValaIle 256
QY 371 GCTCCT---GGCGTTTCTCCTAGCCCTATCTACTATATCCGATGCTGATCCTGATCAA 427
D 257 SerProLeuGlyIleIleAspThrAlaValCysIleArgGlu----- 270
QY 428 CTTGGAATAAATCGGACGAGCTGAAATGCAATTTGATTTATTTAAAGATTAAAGA 487
D 271 -----IleGluLysIleLeuAsnAsnLeuSerLeuAsnGlyTyrGlnSerSerLeuPro 288
QY 488 CACGATTTTAAATA-----AAGTCTTGATGACGATTTTCCAAAATAAACAATAAT 541

```

Db      289  GluGlyHisLysLeuThrAsnGluGluLeuAsnGlyLeuSerGlnLysAsnAsnGlu--- 307
QY      542  ATTGATACTATAAGTAATATTACTAGAACTGGGTACTTATTAGATGATTCTTATCGT 601
Db      308  -----ProSerLeuProGluGluHisArgGlnLeuAspGlyAlaLeuAsp 322
QY      602  ATGATGGAAACAAATACATAATATC---AATAAGTTGTCTAAAGAATT----- 648
Db      323  SerLysSerGlnLysThrTyrAsnPheGluAsnLysTyrIleLysGlnGlnThrArgPhe 342
QY      649  -----GCAAACTGTTAGCCAAACCAATCAGCATTGTCTATGTGTAGTGCAACCAAA 699
Db      343  IleSerGlnAlaAlaTyrPheSerArgSerIleAspCysLeuAsnLeuThrAlaLysLys 362

```

Search completed: May 13, 2004, 08:26:09
Job time : 28.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:21:52 ; Search time 26.5 seconds
(without alignments)
8479.374 Million cell updates/sec

Title: US-10-030-529A-1

Perfect score: 2048

Sequence: 1 ataataacgtcaatgacatt.....aagccggttaaagtgcgac 1168

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp

-Q=/cg2_1/USPTO.spool.P/US10030529/runat 13052004 081453 2632/app_query.fasta_1.1351

-DB=PIR_78 -QFMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THRT=SCORE=1 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptio -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10030529 @CGN_1_1_38 @runat 13052004 081453 2632 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	10.6	364	2 A81019	adhesin/invasin, p
2	186.5	9.1	455	2 S04912	yopA protein - Yer
3	186	9.1	422	2 S04911	yopA protein - Yer
4	182	8.9	434	2 S04534	invasin precursor
5	144	7.2	403	1 H64142	hypothetical prote
6	132	6.4	155	2 AH3277	immunoglobulin-bin
7	130	6.5	227	1 H65039	hypothetical prote
8	129	6.5	227	2 G85907	probable transport
9	129	6.5	398	2 C91063	hypothetical prote
10	129	6.5	413	2 AC0834	hypothetical prote
11	119.5	5.8	658	2 AE0110	probable surface p
12	112.5	5.5	364	2 AE0169	probable exported
13	109.5	5.3	338	2 D90697	adhesin/invasin-li
14	109.5	5.3	338	2 G85547	adhesin/invasin-li

C 15	107	5.4	427	2 AF0400	probable membrane
C 16	104	5.2	426	2 G82309	probable hemolysin
C 17	102.5	5.0	498	2 H69623	flagellar hook-ass
C 18	102.5	5.0	622	2 AF0169	probable exported
C 19	99.5	4.9	466	2 H98045	hypothetical prote
C 20	99.5	4.9	831	2 S76235	hypothetical prote
C 21	99.5	4.9	1695	2 A56921	kinesin family pro
C 22	99	4.8	795	1 D64226	endopeptidase La (
C 23	98.5	4.8	466	2 A95179	hypothetical prote
C 24	97.5	4.8	729	2 T45780	sugar transporter-
C 25	95	4.6	276	2 S05856	gene Atrase6 intro
C 26	95	4.6	2214	1 A48548	genome polyprotein
C 27	94.5	4.6	301	2 S44255	hypothetical prote
C 28	94.5	4.6	1036	2 B69368	hypothetical prote
C 29	93.5	4.6	163	2 T28380	ORF MSV219 hypothe
C 30	92.5	4.5	588	2 T22888	hypothetical prote
C 31	92.5	4.5	940	2 S19702	fibronectin-bindin
C 32	92	4.5	316	2 G81045	pilin gene inverti
C 33	92	4.5	335	2 AF1125	phosphate transpor
C 34	92	4.5	1338	2 T02206	hypothetical prote
C 35	91	4.4	659	2 AB3031	conserved hypochet
C 36	91	4.4	765	2 E98254	hypothetical prote
C 37	91	4.4	2715	2 T13049	eyelid - fruit fly
C 38	90.5	4.4	340	2 H81348	probable periplasm
C 39	90.5	4.4	819	1 H72128	endopeptidase La (
C 40	90.5	4.4	819	2 E86494	Lon ATP-dependent
C 41	90.5	4.4	827	2 AD1311	penicillin-binding
C 42	90.5	4.4	1212	2 A96971	cobalamine-depende
C 43	90.5	4.4	1226	1 S65593	adenosine deaminas
C 44	90	4.4	870	2 S65158	PAL1 protein - yea
C 45	90	4.4	1328	1 S04273	retrovirus-related

ALIGNMENTS

RESULT 1

A81019

adhesin/invasin, probable NMB1994 [imported] - Neisseria meningitidis (strain MC58 serog

C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C/Accession: A81019

R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A/Reference number: A81000; MUID:20175755; PMID:10710307

A/Accession: A81019

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-364 <TET>

A/Cross-references: GB:AE002548; GB:AE002098; NID:g7227249; PIDN:AAF42321.1; PID:g7227275

A/Experimental source: serogroup B, strain MC58

C/Genetics:

A/Gene: NMB1994

Alignment Scores:

Pred. No.:	1.47e-10	Length:	364
Score:	217.00	Matches:	52
Percent Similarity:	53.79%	Conservative:	19
Best Local Similarity:	39.39%	Mismatches:	37
Query Match:	10.60%	Indels:	24
DB:	2	Gaps:	3

US-10-030-529A-1 (1-1168) x A81019 (1-364)

Qy 479 GATTTAAGACACGATTTTAAATTAAGTTCTTGATGCAGCATTTTCCAAAATAAACAA 538

Db 256 AspileyalaalaspilealathrAsnLysAlaAsp-----lleAlatysAsnSerAla 273

Qy 539 AATATTGATTAAGTAAATATTACTAGAACTGGGTACTTATTATTAGATGATTCCTAT 598

```
Db 274 ArgileAspSerLeuAspLys----- 280
Qy 599 CGTATGATGGACAAAATACACATAATATCAATAAGTTGTCTAAAGATTTCGAACATGGT 658
Db 281 -----AsnValAlaLeuLeuArgLysGluThrArgGlnGly 292
Qy 659 TTAGCCAAACCATCAGCATTTCTATGTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
Db 293 LeuAlaGluGlnAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 312
Qy 719 GTTCTGTCGGGTAGGAGTTATAGACATAAACTGCATTAGCCATTGGTGGCTCA 778
Db 313 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 332
Qy 779 CGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACCAATGGC--- 835
Db 333 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 352
Qy 836 GGCATGCTTATGGTCTCTCTGTGGTTTATGAATTC 871
Db 353 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 364

RESULT 2
S04912
YopA protein - Yersinia enterocolitica plasmid pYV6471/76
C;Species: Yersinia enterocolitica
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S04912
R;Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A;Title: Analysis of the YopA gene encoding the YopI virulence determinants of Yersinia
A;Reference number: S04910; MUID:89343638; PMID:2761389
A;Accession: S04912
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-455 <SKU>
A;Cross-references: EMBL:X13882; NID:g48606; PIDN:CAA32086.1; PID:g48607
C;Genetics:
A;Genome: plasmid pYV6471/76

Alignment Scores:
Pred. No.: 5.86e-08 Length: 455
Score: 186.50 Matches: 47
Percent Similarity: 52.1% Conservative: 25
Best Local Similarity: 34.0% Mismatches: 57
Query Match: 9.11% Indels: 9
DB: 2 Gaps: 4

US-10-030-529A-1 (1-1168) x S04912 (1-455)
Qy 458 AATTGTGATAGTTATTTTAAACGATTTAAAGACGATTTTAAATTAAGTTCTTGATGCA 517
Db 327 AsnValTyrAla-----AspSerLysSerHisThrLeuLysThrAlaAsnSer 343
Qy 518 CGTATTTCCAAAAATAAACAATAATGATACATAAGTAAATATTTACTAGAACTGGGT 577
Db 344 TyrThrAspValThrValSerAsn-----SerThrLysLysAlaIleArgGluSerAsn 361
Qy 578 ACTTATTTAGATGATCTTATGATGATGATGGAACAAATACACATAATATCAATAAGTTG 637
Db 362 GlnTyrThrAspHisLysPheArgGlnLeuAsp-----AsnArgLeuAspLysLeu 378
Qy 638 TCTAAGAATGTCAAACTGGTTTACCGCAACCAATCAGCATTTCTATGTTAGTGCACCA 697
Db 379 AspThrArgValAspLysGlyLeuAlaSerSerAlaAlaLeuAsnSerLeuPheGlnPro 398
Qy 698 AATGTTGTAGGCAAAACGAGCTTCTCTGCTGGGTAGGAGTTATAGAGATAAAACTGCA 757
Db 399 TyrGlyValGlyLysValAsnPheThrAlaGlyValGlyGlyTyrArgSerSerGlnAla 418
Qy 758 TTAGCATTTGGTGGCTCGCATCTACGCATTTACTGATCGCTTTACCGCTAAAGCGGTGATCG 817
Db 419 LeuAlaIleGlySerGlyTyrArgValAsnValAsnValAlaLeuLysAlaGlyValAla 438
```

```
Qy 818 TTCAATACCTACATGCGGCATGCTCTTATGGTCTTCTGTTGTTATGAATTC 871
Db 439 Tyr---AlaGlySerSerAspValMetTyrAsnAlaSerPheAsnIleGluTrp 455

RESULT 3
S04911
YopA protein - Yersinia enterocolitica virulence plasmid pYV8081
C;Species: Yersinia enterocolitica
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S04911
R;Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A;Title: Analysis of the YopA gene encoding the YopI virulence determinants of Yersinia
A;Reference number: S04910; MUID:89343638; PMID:2761389
A;Accession: S04911
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-422 <SKU>
A;Cross-references: EMBL:X13881; NID:g48609; PIDN:CAA32085.1; PID:g48610
A;Note: the authors translated the codon CCA for residue 282 as Thr

Alignment Scores:
Pred. No.: 6.45e-08 Length: 422
Score: 186.00 Matches: 44
Percent Similarity: 51.56% Conservative: 22
Best Local Similarity: 34.38% Mismatches: 46
Query Match: 9.08% Indels: 16
DB: 2 Gaps: 3

US-10-030-529A-1 (1-1168) x S04911 (1-422)
Qy 524 TCCAAAAATAACAAAATATTTGATACCTAATAAGTAATAT----- 562
Db 299 SerLysSerSerHisThrLeuGlnThrAlaAsnSerTyrThrAspValThrValSerAsn 318
Qy 563 -----TTACTAGAACTGGTACTTATTAGATGATCTTATTCGTTATGATG 607
Db 319 SerThrLysLysAlaIleArgGluSerAsnGlnTyrThrAspHisLysPheArgGlnLeu 338
Qy 608 GAACAAAATACACATAATATCAATAAGTTGCTCTAAAGAAATTCGAACTGGTTAGCCAAC 667
Db 339 Asp-----AsnArgLeuAspLysLeuAspThrArgValAspLysGlyLeuAlaSer 355
Qy 668 CAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGTTTCTGCT 727
Db 356 SerAlaAlaLeuAsnSerLeuPheGlnProTyrGlyValGlyLysValAsnPheThrAla 375
Qy 728 GCGTAGGAGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCACGCAATCTACT 787
Db 376 GlyValGlyGlyTyrArgSerSerGlnAlaLeuAlaIleGlySerGlyTyrArgValAsn 395
Qy 788 GATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACATGCGGCATCTCTAT 847
Db 396 GluSerValAlaLeuLysAlaGlyValAlaTyr---AlaGlySerSerAspValMetTyr 414
Qy 848 GGTGCTTCTGTTGTTATGAATTC 871
Db 415 AsnAlaSerPheAsnIleGluTrp 422

RESULT 4
S04534
invasin precursor - Yersinia pseudotuberculosis plasmid pIB1
C;Species: Yersinia pseudotuberculosis
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S04534; S04910
R;Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.
Nature 334, 522-525, 1988
A;Title: Increased virulence of Yersinia pseudotuberculosis by two independent mutations
A;Reference number: S04534; MUID:88302441; PMID:3043229
A;Accession: S04534
A;Status: preliminary
A;Molecule type: DNA
```

Pred. No.:	0.000242	Length:	403
Score:	144.00	Matches:	29
Percent Similarity:	81.40%	Conservative:	6
Best Local Similarity:	67.44%	Mismatches:	8
Query Match:	7.22%	Indels:	0
DB:	1	Gaps:	0
US-10-030-529A-1 (1-1168) x H64142 (1-403)			
Qy	1165	CGCACCTTTAAACGGCTTAATTTTAGAACATTTAGAAAAAATTCGATAGAGGTACCCAA	1106
Db	358	ArgThrPheAsnGlyLeuIleLeuGluHisLeuGluGluIleProAspGluGlyThrIle	377
Qy	1105	TTTAGGCTAAACCAACTTAATATATACCGTTTTHAGAAGTCGCGGATATATGTTGTAAGAAA	1046
Db	378	CysGluIleAspGlyLeuLeuIleThrIleLeuGluValGlyAspAsnMetIleLysGln	397
Qy	1045	GTGNAAGTA	1037
Db	398	AlaLysVal	400
RESULT 6			
AH3277			
immunoglobulin-binding protein eibE [imported] - Brucella melitensis (strain 161			
C;Species: Brucella melitensis			
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002			
C;Accession: AH3277			
R;DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;			
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.			
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002			
A;Title: The genome sequence of the facultative intracellular pathogen Brucella			
A;Reference number: AD3252; PMID:11756688			
A;Accession: AH3277			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-155 <KUR>			
A;Cross-references: GB:AE008917; PIDN:AAL51387.1; PID:g17982090; GSPDB:GN00190			
A;Experimental source: strain 16M			
C;Genetics:			
A;Gene: BMEI0205			
A;Map position: I			
Alignment Scores:			
Pred. No.:	0.00245	Length:	155
Score:	132.00	Matches:	40
Percent Similarity:	45.52%	Conservative:	21
Best Local Similarity:	29.85%	Mismatches:	35
Query Match:	6.45%	Indels:	38
DB:	2	Gaps:	5
US-10-030-529A-1 (1-1168) x AH3277 (1-155)			
Qy	521	ATTTCACAAATTAACAAATATTGTACTATTAAGTAATATTTACTAGACTGGTACT	580
Db	32	IleGlyGlnAsnArgThrSerIleSerThrAsnAlaLysGlyValala	47
Qy	581	TATTAGATCTTATCGTATGATGGAACAAATACA	619
Db	48	-----AspAsnArgAlaAlaIleArgGlnAsnSerAlaAlaIleSerAlaLeuGly	64
Qy	620	-----CATATATCAATAAGTTCTCTAAAGAAATTGCAAACTGGT	658
Db	65	GlnArgValAspGlyLeuGlnGlyGlnIleAsnSerAlaArgLysGluAlaArgAlaGly	84
Qy	659	TTAGCCAAACCATCAGCATTTCTATTGTAGTGCAACCAAAATGGTGTAGGCAAAACGAGC	718
Db	85	AlaAlaAsnAlaAlaAlaLeuSerGlyLeuArgTyrAspAsnArgProGlyLysValSer	104
Qy	719	GTWTTCTGCTCCGTAGAGGTTATAGATATAAACTGCATTAGCCATTGGTTCGCGCTCA	778
Db	105	IleAlaThrGlyValGlyGlyPheLysGlySerThrAlaLeuAlaGlyIleGly	123
Qy	779	CGCATTTACTGATCGCTTTTACCGCTAAA	817


```

124  -----TyrThrSerLysAsnGluAlaArgTyrAsnValSerValAla 138
      :::::|||||
Qy  818  TTCAAT-----ACCTACAATGGCGCATGCTTAT 847
      ::|||
Db  139  TyrAsnGluAlaGlyThrSerTrpAsnAlaGlyAlaSerPhe 152
      :::::|||||

RESULT 7
H65039
hypothetical protein in grpE 3' region - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: H65039
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H65039
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-227 <BLAT>
A;Cross-references: GB:AB000347; GB:U00096; NID:92367142; PIDN:AAC75662.1; PID:gl788966;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yfjD
C;Superfamily: yfjD protein

Alignment Scores:
Pred. No.: 0.00368 Length: 227
Score: 130.00 Matches: 24
Percent Similarity: 81.40% Conservatives: 11
Best Local Similarity: 55.81% Mismatches: 8
Query Match: 6.52% Indels: 0
DB: 1 Gaps: 0

US-10-030-529A-1 (1-1168) x H65039 (1-227)

Qy  1165  CGCACCTTTACGGCTTAATTTAGAACATTTAGAAAAAATTCGATAGAGGTACCCAA 1106
      |||||
Db  173  ArgThrValAsnGlyValIleLeuGluAlaLeuGluProValGlyThrArg 192
      |||||

Qy  1105  TTTAGGCTAAACCAACTTAATATACCGTTTTCAGAGTGGCGGATAATATGGTGAAGAAA 1046
      |||||
Db  193  ValArgIleGlyGluTyrAspIleAspIleLeuAspValGlnAspAsnMetIleLysGln 212
      |||||

Qy  1045  GTGAAAGTA 1037
      |||||
Db  213  ValLysVal 215
      |||||

RESULT 8
G85907
probable transport protein yfjD [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: G85907
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85907
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <STO>
A;Cross-references: GB:AB005174; NID:912517037; PIDN:AAG57723.1; GSPDB:GN00145; UMGF:Z39
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yfjD
C;Superfamily: yfjD protein

Alignment Scores:
Pred. No.: 0.00448 Length: 227
```

```

Score: 129.00 Matches: 24
Percent Similarity: 81.40% Conservatives: 11
Best Local Similarity: 55.81% Mismatches: 8
Query Match: 6.47% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x G85907 (1-227)

Qy  1165  CGCACCTTTACGGCTTAATTTAGAACATTTAGAAAAAATTCGATAGAGGTACCCAA 1106
      |||||
Db  173  ArgThrValAsnGlyValIleLeuGluAlaLeuGluProValGlyThrArg 192
      |||||

Qy  1105  TTTAGGCTAAACCAACTTAATATACCGTTTTCAGAGTGGCGGATAATATGGTGAAGAAA 1046
      |||||
Db  193  ValArgIleGlyGluTyrAspIleAspIleLeuAspValGlnAspAsnMetIleLysGln 212
      |||||

Qy  1045  GTGAAAGTA 1037
      |||||
Db  213  ValLysVal 215
      |||||

RESULT 9
C91063
hypothetical protein ECs3475 [imported] - Escherichia coli (strain O157:H7, substrain R1
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
C;Accession: C91063
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C91063
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA036898.1; PID:gl3362946; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain R1MD 0509952
C;Genetics:
C;Superfamily: hypothetical protein HI0107

Alignment Scores:
Pred. No.: 0.00458 Length: 398
Score: 129.00 Matches: 24
Percent Similarity: 81.40% Conservatives: 11
Best Local Similarity: 55.81% Mismatches: 8
Query Match: 6.47% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x C91063 (1-398)

Qy  1165  CGCACCTTTACGGCTTAATTTAGAACATTTAGAAAAAATTCGATAGAGGTACCCAA 1106
      |||||
Db  344  ArgThrValAsnGlyValIleLeuGluAlaLeuGluProValGlyThrArg 363
      |||||

Qy  1105  TTTAGGCTAAACCAACTTAATATACCGTTTTCAGAGTGGCGGATAATATGGTGAAGAAA 1046
      |||||
Db  364  ValArgIleGlyGluTyrAspIleAspIleLeuAspValGlnAspAsnMetIleLysGln 383
      |||||

Qy  1045  GTGAAAGTA 1037
      |||||
Db  384  ValLysVal 386
      |||||

RESULT 10
AC0834
probable membrane protein corB [imported] - Salmonella enterica subsp. enterica serovar
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0834
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
```


Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AC0834

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-413 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05858.1; PID:g16503833; GSPDB:GN00176

C:Genetics:

A:Gene: corB

C:Superfamily: hypothetical protein HI0107

Alignment Scores:

Pred. No.: 0.00459 Length: 413

Score: 129.00 Matches: 24

Percent Similarity: 81.40% Conservative: 11

Best Local Similarity: 55.81% Mismatches: 8

Query Match: 6.47% Indels: 0

DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x AC0834 (1-413)

Qy 1165 CGCACCTTAACGGCTTAATTTAGAACATTTAGAAAAATTCGATAGAAAGTACCCAA 1106

Db 359 ArgThrValAsnGlyValIleLeuGluAlaLeuGluGluIleProValAlaGlyThrArg 378

Qy 1105 TTTAGGCTTAACCACTTAATTAATACCGTTTGTAGAGTGGCGGATAATATGTTGAAAAA 1046

Db 379 ValArgIleGluGlnTyrAspIleAspValGlnGluAsnMetIleLeuGln 398

Qy 1045 GTGAAGTA 1037

Db 399 ValIysVal 401

RESULT 11

AH0110

probable surface protein (partial) YPO0902 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AH0110

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0110

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-658 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175

C:Genetics:

A:Gene: YPO0902

Alignment Scores:

Pred. No.: 0.0301 Length: 658

Score: 119.50 Matches: 31

Percent Similarity: 49.56% Conservative: 25

Best Local Similarity: 27.43% Mismatches: 50

Query Match: 5.83% Indels: 7

DB: 2 Gaps: 4

US-10-030-529A-1 (1-1168) x AH0110 (1-658)

Qy 539 AATATTGATCTAATAAGTAATATTACTAGAACTGGGTACTTATTATGATGATCTTAT 598

Db 551 AsnPheAspGlnLeuLysSerIleSerAsnGlnThrAsnAlaTyrThrAsnGlnArgTyr 570

Qy 599 CGTATGTGGACAAATACACATAATCAATCAATGTTGCTAAAGAAATTCGAACTGGT 658

Db 571 SerGluLeuLysGln-----AspLeuArgLysGlnAsnSerValLeuSerAlaGly 587

Qy 659 TTAGCAACCAATCAGCATTCGTCTATGTTAGTCAACCAATGGTGTAGCAAAACGAGC 718

Db 588 IleAlaSerAlaMetSerMetAlaSerLeuThrGlnProTyrThrSerGlySerSerMet 607

Qy 719 GTTTCGTGCGGTAGGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCA 778

Db 608 ThrThrIleGlyAlaAlaSerTyrArgGlyGlnSerAlaLeuSerLeuGlyValSerSer 627

Qy 779 CGCATTTACTGAT-----CGCTTTACCGCTAAAGCGGTGTAGCGTTTCAATCACTCAAT 832

Db 628 ---IleSerAspSerGlyArgTyrValSerLysLeuGlnAlaSerSerAsnThr---Gln 645

Qy 833 GCGCGCATGCTTATGGTCTCTGTTGGTTGATGAATTC 871

Db 646 GlyAspPheGlyIleGlyValGlyValGlyTyrGlnTrp 658

RESULT 12

AE0169

probable exported protein YPO1387 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AE0169

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AE0169

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC90216.1; PID:g15979436; GSPDB:GN00175

C:Genetics:

A:Gene: YPO1387

Alignment Scores:

Pred. No.: 0.116 Length: 364

Score: 112.50 Matches: 40

Percent Similarity: 43.62% Conservative: 25

Best Local Similarity: 26.85% Mismatches: 69

Query Match: 5.45% Indels: 15

DB: 2 Gaps: 6

US-10-030-529A-1 (1-1168) x AE0169 (1-364)

Qy 431 GGAATAATCGCAGCAGCTCAAAATTCGAATTTGATAGTATTATTAAACGATTTAAGACAC 490

Db 226 GlyLysSerLysGluThrLeuAsn---AsnThrTyrAspTyrVal----- 239

Qy 491 GATTTTAAATTAATA-----GTTCTTGATCGACGATTTTCCAAAAATAAAATAATATT 544

Db 240 AspSerLysValSerSerIleValTyrAspValAsnSerTyrThrAspLysThrValAsn 259

Qy 545 GATCTATAAGTAAATATTATTACTAGAACTGGGTACTTATTAGATGATCTTATCGTATG 604

Db 260 ThrAlaPheGluThrSerLeuSerAspAlaLysSerTyrValAspAspLysTyr----- 277

Qy 605 ATGGAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTCGAACTGGTTAGCC 664

Db 278 ---AsnGlnLeuSerAspLysValAsnLysAsnPheAsnLysThrAsnAlaGlyIleSer 296

Qy 665 AACCAATCAGCATTCGTCTATGTTAGTCAACCAATGGTGTAGCAAAACGAGCGTTTCT 724

Db 297 GlyAlaMetAlaMetSerGlyIleProGlnLysPheGlyTyrGluLys---SerPheGly 315

Qy 725 GCTGCGGTAGAGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTACGCCATT 784

Db 316 MetAlaIleGlyAlaTyrArgGlyGlnSerAlaLeuAlaValGlyAspTyrAsnIle 335

Qy 785 ACTGATCGCTTACCGCTAAAGCGGTGTAGCGTTTCAATCACTCAATGGCGGATGTCT 844

Db 336 AsnHisLysThrIleThrArgValAsnValSerAlaAspThrGluGly----- 352

Search completed: May 13, 2004, 08:29:37
Job time : 32.5 secs

```
Db      185 ThrAspValAlaValGlnGlnGlnAsnGluArgThrGlnTyrAspLysGlnMetGlnSer 204
Qy      539 -----AATATTGATCTATAAGTAAATATTACTA 568
Db      205 LeuAlaGlnGluSerAlaGlnAlaHisGluGlnIleAspSerLeuSerGlnAspValThr 224
Qy      569 GAATGGGTACTTATTAGATTCTTATCGTATGGAACAATAACACATATAATATC 628
Db      225 GlnThrHisGlnGlnLeuThrAsnThrGlnLysArgValAlaAspAsnSerGlnGlnIle 244
Qy      629 AATAAGTTGTCT-----AAAGAA 646
Db      245 AsnThrLeuAsnAsnHisPheSerSerLeuLysAsnGluValAspAsnArgLysGlu 264
Qy      647 TTGCAAACTGGTTAGCAACCAATGACATGCTATGTTAGTGCAACCAATGGTGTA 706
Db      265 AlaAsnAlaGlyThrAla-----SerAlaIleAlaIleAlaSerGlnProGln----- 280
Qy      707 GGCRAAAGCAGC-----GTTCTGCTCGGTAGGAGGTTATAGAGTAAACT 754
Db      281 ValLysThrGlyAspValMetMetValSerAlaGlyAlaGlyThrPheAsnGlyGluSer 300
Qy      755 GCATTAGCCATTGGTGTGCGCTCACGCATTACTGCTTACCGCTAAAGCGGGTGA 814
Db      301 AlaValSerValGlyThrSerPheAsnAlaGlyThrHisThrValLeuLysAlaGlyIle 320
Qy      815 GCGTTCAATCACTACATGCGGGCATGCTTATCGTCTCTGTTGTTATGAATTC 871
Db      321 SerAlaAspThr---GlnSerAspPheGlyAlaGlyValGlyValGlyTyrSerPhe 338
```

RESULT 15

AF0400
probable membrane protein YPO3298 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF0400
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413, PMID:11586360
A:Accession: AF0400
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92530.1; PID:gl5981228; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3298
C:Superfamily: hypothetical protein HI0107

Alignment Scores:
Pred. No.: 0.343 Length: 427
Score: 107.00 Matches: 21
Percent Similarity: 67.44% Conservative: 8
Best Local Similarity: 48.84% Mismatches: 14
Query Match: 5.36% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x AF0400 (1-427)

```
Qy      1165 CGCACCTTTTACGGCTTAATTTTAGAATAATTTAGAAAAAATTCGATAGAGTACCCAA 1106
Db      374 ArgThrIleAsnGlyMetLeuLeuGluGluGluGluGluIleProGlnAlaHisIleHis 393
Qy      1105 TTTAGGCTAAACCAACTTAATATATACCGTTTGAAGTGGCGGATATATATCGTGAATAAAA 1046
Db      394 ValArgLeuGlyAsnTyrGlnIleGluValLeuAspValGlnGlnAsnMetIleLysArg 413
Qy      1045 GTGAAGTA 1037
Db      414 Vallysile 416
```

This Page Blank (uspto)